DOCUMENT RESUME

ED 067 104 LI 003 849

TITLE	Information Retrieval.
INSTITUTION	Air Force, Washington, D.C.; National Archives and Records Service (GSA), Washington, D.C. Office of the National Archives.
REPORT NO	FPMR-101-11.3
PUB DATE	72
NOTE	135p.; (19 References); Third of the series "Managing Information Retrieval"
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Automatic Indexing; *Coordinate Indexes; Guidelines; Indexing; *Information Retrieval; *Information Systems; Management; *Microforms

ABSTRACT

It is the purpose of this handbook to provide managers and their assistants with guidelines for determining where new information processing systems (dissemination, storage and retrieval) might profitably be employed in Government offices and with criteria for selecting the right methods and equipment. While the main objective is to encourage greater use of modern information retrieval techniques, the guidelines should also help prevent the installation of ill-advised or unprofitable systems. For those offices that have already installed modern information retrieval systems, the handbook may prove helpful in analyzing and evaluating existing system performance or in revising an ineffective system. This handbook is intended primarily for the use of management analysts, systems personnel, middle management, and any others who may be directly involved in conducting information retrieval studies or in designing and installing an information retrieval system. (Author/SJ)

ED 067104

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

FPMR 101—11.8

RECORDS MANAGEMENT HANDBOOK

Managing Information Retrieval

INFORMATION RETRIEVAL

1972

GENERAL SERVICES ADMINISTRATION
NATIONAL ARCHIVES AND RECORDS SERVICE
OFFICE OF RECORDS MANAGEMENT

Federal Stock Number
7610-042-8782

1

RECORDS MANAGEMENT HANDBOOKS are developed by the National Archives and Records Service as technical guides to reducing and simplifying paperwork.

RECORDS MANAGEMENT HANDBOOKS:

Managing correspondence: *Plain Letters*	1955	47 p.
Managing correspondence: *Form Letters*	1954	33 p.
Managing correspondence: *Guide Letters*	1955	23 p.
Managing directives: *Communicating Policy and Procedure*	1967	62 p.
Managing forms: *Forms Analysis*	1960	62 p.
Managing forms: *Forms Design*	1960	89 p.
Managing forms: *Forms Management*	1969	34 p.
Managing mail: *Managing the Mail*	1971	94 p.
Managing current files: *Files Operations*	1964	76 p.
Managing current files: *File Stations*	1967	52 p.
Managing current files: *Subject Filing*	1966	40 p.
Managing information retrieval: *Information Retrieval*	1972	132 p.
Managing information retrieval: *Information Retrieval Systems*	1970	150 p.
Managing information retrieval: *Microform Retrieval Equipment Guide*	1970	64 p.
Managing emergency preparedness files: *Federal Vital Records Program*	1968	16 p.
Managing noncurrent files: *Applying Records Schedules*	1961	23 p.
Managing noncurrent files: *Federal Records Centers*	1967	39 p.
Mechanizing paperwork: *Source Data Automation*	1965	78 p.
Mechanizing paperwork: *Source Data Automation Equipment Guide*	1970	152 p.
Mechanizing paperwork: *Source Data Automation Systems*	1963	183 p.
General: *Bibliography for Records Managers*	1965	58 p.
General: *Copying Equipment*	1966	82 p.

ED 067104

FOREWORD

Management at every level is being subjected to increasing pressure to automate the files of the office—to adopt new, nonconventional methods and equipment to improve the dissemination, storage, and retrieval of information. Professional journals, trade magazines, and agency publications are constantly reporting how management is solving its information problems through the use of these new systems. But today's manager knows that the new systems usually represent a sizable investment, and he is also aware that the investment has not always paid off.

It is the purpose of this handbook to provide the manager and those who assist him with guidelines for determining where these new systems might profitably be employed in Government offices and with criteria for selecting the right methods and equipment. While the main objective is to encourage greater use of modern information retrieval techniques, the guidelines should also help prevent the installation of ill-advised or unprofitable systems. For those offices that have already installed modern information retrieval systems, the handbook may prove helpful in analyzing and evaluating existing system performance or in revising an ineffective system.

This handbook is intended primarily for the use of management analysts, systems personnel, middle management, and any others who may be directly involved in conducting information retrieval studies or in designing and installing an information retrieval system.

Although this handbook is issued as one of a series of Records Management Handbooks produced by the National Archives and Records Service, General Services Administration (GSA), the United States Air Force shared in its development. It was produced under a contract jointly funded and administered by the Air Force and GSA.

CONTENTS

Page

I. WHY NEW INFORMATION RETRIEVAL SYSTEMS ARE NEEDED
- What Is Information Retrieval? 1
- Summary of Conventional Methods 1
- Summary of Nonconventional Methods ... 3
- Limitations and Advantages of Conventional Methods 5
- Advantages and Limitations of Nonconventional Methods 9
- Coordinating Indexing—Key to Many Nonconventional Systems 10

II. HOW COORDINATE INDEXING SYSTEMS WORK
- Principles of Coordinate Indexing 11
- Types of Indexing Terms 12
- Index File Arrangements 13
- Major Advantages of Coordinate Indexing Systems 13

III. MICROFORM SYSTEMS
- How Microforms Help Solve Typical Information Problems 14
- Prerequisites for a Successful Microform System 18
- Types of Microfilm and Cameras 18
- Factors Affecting the Choice of the Type of Microform System 21
- Types of Microform Systems 22
- Microform-Computer Combinations 31
- Special Considerations 32

IV. MANUAL NONCONVENTIONAL INDEXING SYSTEMS
- Types of Situations Where Nonconventional Indexing Systems Are Used 33
- Prerequisites For a Successful Manual Nonconventional Indexing System 33
- Factors Affecting the Choice of the Type of Manual Nonconventional Indexing System 34
- Types of Manual Nonconventional Systems 35
- Special Considerations 40

V. NONCONVENTIONAL MACHINE INDEXING RETRIEVAL SYSTEMS
- Types of Situations Where Machine Indexing and Retrieval Systems Apply 41

Page

- Prerequisites for a Successful Machine Indexing or Retrieval System 41
- Factors Affecting the Choice of the Type of a Machine Indexing or Retrieval System 42
- Types of Machine Indexing and Retrieval Systems 42
- Other Machine Indexing and Retrieval Systems 50

VI. HOW TO DECIDE IF A NEW SYSTEM IS NEEDED
- The Preliminary Survey 51
- Where to Look 51
- Examining User Needs 51
- Fact-Gathering Forms 53
- Decision Tables 53
- Summary 59

VII. HOW TO DETERMINE SYSTEM REQUIREMENTS
- Data Collection Techniques 60
- Suggested Questionnaires 60
- Data Summarization Techniques 65
- Final Review and Analysis of Findings ... 71
- Users' Briefings 71
- Use of General Analysis Techniques and Tools 71

VIII. SELECTING THE RIGHT METHODS AND EQUIPMENT
- Step 1, Selecting the Applicable Functional Category 72
- Step 2, Selecting the Right Methods and Equipment 73

IX. DESIGNING A COORDINATE INDEX
- Economics of Coordinate Indexes 76
- Steps in Developing a Coordinate Index .. 76
- Staffing 85
- Current Awareness Services 85
- Quality Control 88
- Setting Quality Standards 88
- Conclusion 89

APPENDIX A NONCONVENTIONAL METHODS AND EQUIPMENT GUIDE 90

APPENDIX B INFORMATION RETRIEVAL EQUIPMENT AND SUPPLIES SOURCES 102

APPENDIX C INFORMATION RETRIEVAL—RECOMMENDED PRIMERS AND SELECTED RESEARCH SOURCES 108

APPENDIX D FORMS FOR EVALUATING A POTENTIAL INFORMATION RETRIEVAL APPLICATION 111

APPENDIX E SAMPLE DIRECTIVE (AIR FORCE) COVERING DOCUMENT MINIATURIZATION SYSTEMS 113

I WHY NEW INFORMATION RETRIEVAL SYSTEMS ARE NEEDED

Conventional methods for storing and retrieving information have been doing an effective information handling job for some 50 years, and in many situations today are still the best answer. However, during and since World War II more and more people have been questioning these conventional methods and looking for new and better ways to satisfy their information needs. The three main reasons for this exploratory research have been the information explosion, the trend toward a much higher degree of specialization in all technical fields, and the advent of the new technologies of electronic data processing and document miniaturization.

The information explosion is now overtaxing conventional methods and equipment for indexing and storing the thousands of new documents being prepared each year. The trend toward greater specialization is resulting in preparation of documents that deal with increasingly narrow aspects of subject topics. New classes of information are constantly being formed by the emergence of interdisciplinary specialists. Conventional methods for classifying and indexing information are frequently not well suited to meet the demands for greater specificity in organizing and retrieving information nor the need to manipulate information freely.

Information specialists in the scientific and technical fields were among the first to apply the electronic computer, microforms, and other nonconventional methods and equipment to solve information retrieval problems. This handbook draws largely on their knowledge and experience.

What Is Information Retrieval?

It is the approach to the problem of information dissemination, storage, and retrieval that is new—nonconventional methods and equipment that have been introduced during the last decade or so. It is this new, nonconventional approach which has become known as "information retrieval." Stated in other ways:

- Information retrieval employs methods and equipment that depart in one way or another from the conventional methods we find in most offices and libraries.
- Information retrieval means there are now available methods and equipment for disseminating, storing, and retrieving information that make it possible, and often quite practical, to do things that no one considered doing before.
- Information retrieval means simply new ways for performing old tasks and is used primarily when conventional methods will no longer suffice.

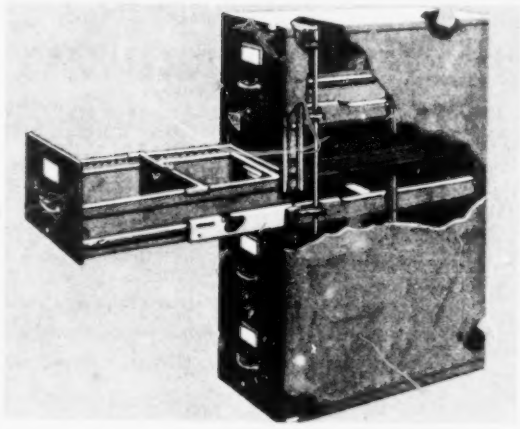
Perhaps one of the best ways to define nonconventional systems is to first explain what is meant by conventional methods and equipment—hence, the things not covered in this handbook. Examples of these conventional methods are shown in figure 1, which includes a standard file cabinet, a reference visible file, a mobile shelf file, a rotary file, and a mechanized file.

Summary of Conventional Methods

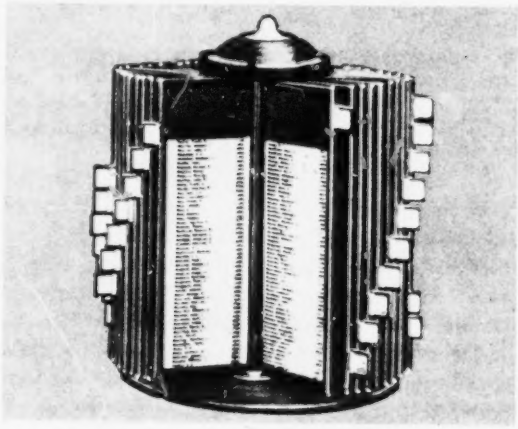
The characteristics of the documents and the methods used in organizing the information in conventional files are as follows:

- The documents are largely in paper form.
- The documents are maintained in a structured file, that is, a file organized and arranged for direct searching according to the filing feature (name, number, subject, etc.) most often known by the user when looking up the information.

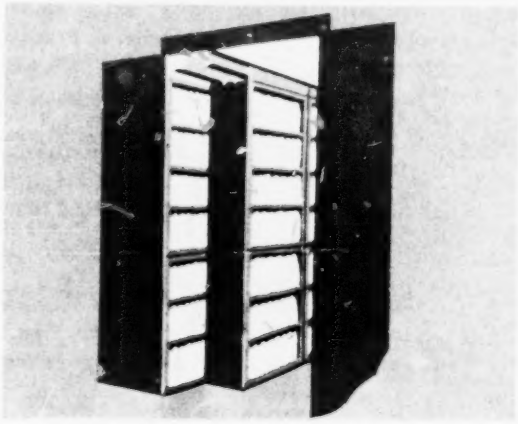
EQUIPMENT FOR CONVENTIONAL FILES



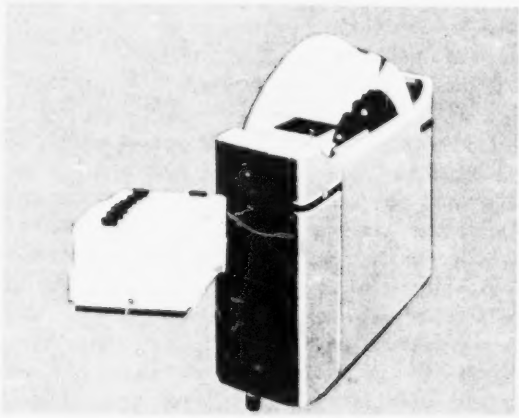
Vertical File Cabinet



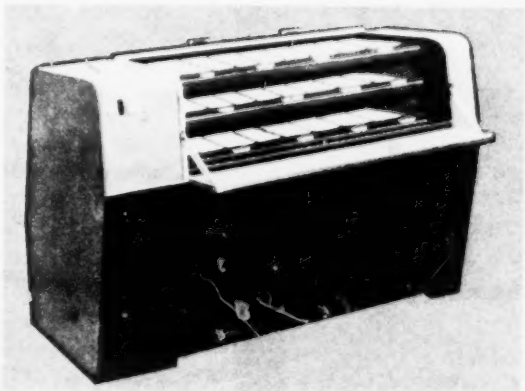
Reference Visible File



Mobile Shelf Files



Rotary File



Mechanized Horizontal File

Figure 1

- If necessary, separate manual index files or finding aids are maintained to help find information when users ask for it on a basis different from that by which the document file is structured.

The success of conventional methods depends largely on the following factors:

- Stability of information and language contained in the documents.
- Simplicity and shortness of the documents.
- Predictability of users' needs and the way in which they will ask for documents.
- Simplicity of users' needs.
- Availability of space close to the users to store the documents.

The following GSA-Records Management Handbooks relate primarily to conventional systems and should be carefully reviewed before any information retrieval study is undertaken:

Files Operations—FSN 7610–985–6973—1964

Subject Filing—FSN 7610–926–2128—1966

File Stations—FSN 7610–926–2129—1967

Summary of Nonconventional Methods

Nonconventional methods for storing and retrieving information have one or more of the following characteristics:

- The information is disseminated and stored in miniaturized form.
- The document file is largely unstructured—the documents are filed by a simple identifier such as an accession number or machine location address.
- The contents of the documents are described in detail by means of a separate, highly manipulative index file, or the entire contents are maintained in machine-readable form.

Figure 2 shows some of the nonconventional methods and equipment employed in modern information retrieval systems and these, with a numbers of others, are described in chapters III, IV, and V of this handbook.

Edge-notched cards. Edge-notched cards have been available for many years and employ a technique that is superior to conventional filing methods in numerous applications.

Optical coincidence cards. The optical coincidence of "peek-a-boo" cards is useful in special applications for organizing and retrieving information.

Microforms. Microfilm was conceived as a recording medium about 100 years ago, and recent developments have made microforms a vital link in solving many of today's information problems.

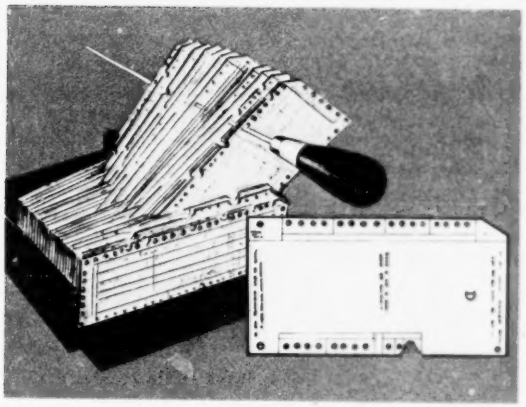
EAM punched cards. EAM (electrical accounting machine) punched cards have been used extensively for processing numerical data, and they can be used readily for storing and retrieving information.

Computers. The most important of the nonconventional tools is the electronic computer, which is playing an increasingly important role in storing and retrieving information.

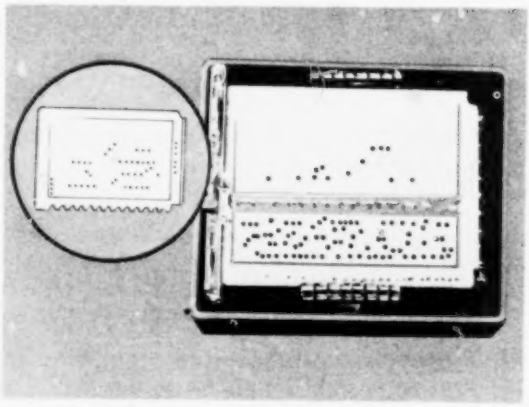
Nonconventional methods can often help when one or more of the following conditions exist:

- Types of information and terminology contained in the document collection are constantly changing.
- Individual documents are lengthy and contain information on a wide variety of subjects or include large quantities of data.
- Users ask for information in a variety of ways and their needs are continuously changing.
- Users' needs are complex in that they require precise information and often must be able to correlate or manipulate it.

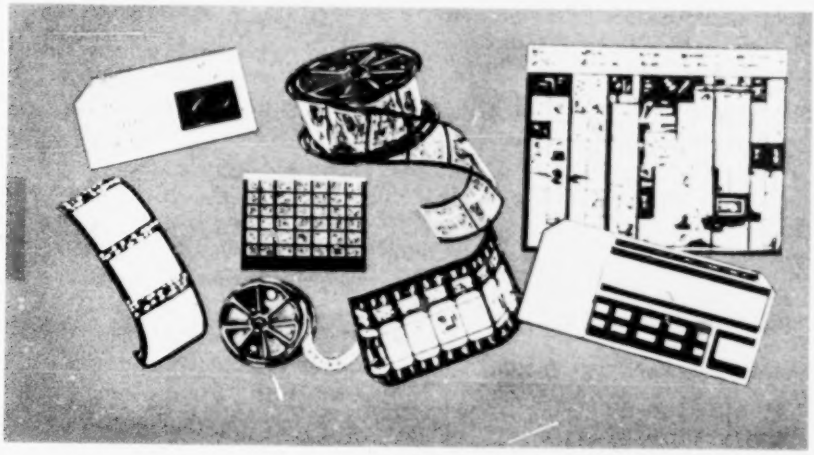
NONCONVENTIONAL METHODS AND EQUIPMENT



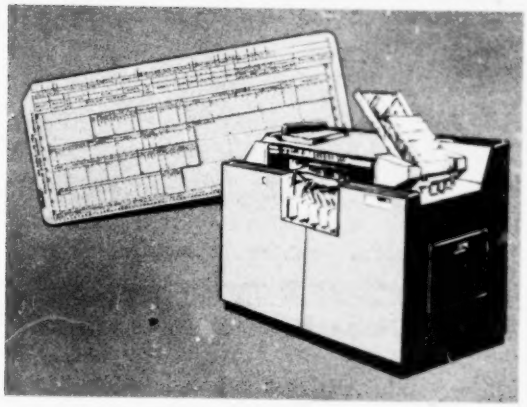
Edge-Notched Cards



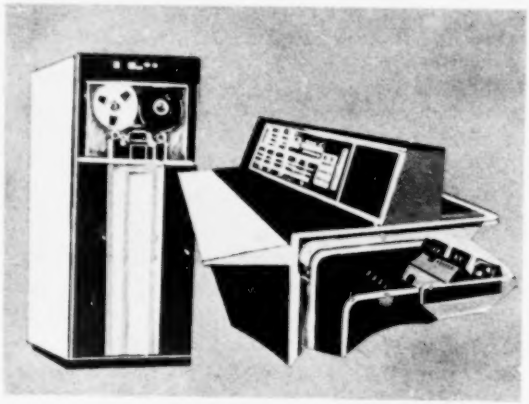
Optical Coincidence Cards



Microforms



EAM Punched Cards



Computers

Figure 2

- Documents need to be reduced in bulk or maintained in multiple sets to facilitate dissemination, storage, and retrieval.

Limitations and Advantages of Conventional Methods

To fully appreciate why nonconventional methods and equipment are needed and where they can best be used, one must first understand the sort of retrieval problems that cannot readily be solved by conventional methods. The three broad types of problems are:

- Location of specific information. Many times today the information the user needs is deeply embedded in a lengthy document—perhaps found in one paragraph of a 50-page research report. If this situation is commonplace and if there are a large number of documents in the collection, retrieval of needed information can be very difficult.

- Location of individual items of data. In some work situations it is frequently necessary to look up individual items of such data as names, numbers, dates, and amounts. The use of such conventional methods as folder files and printed listings for maintaining the data may make this a time-consuming and tiresome chore.

- Conducting coordinate-type searches. In many work situations it is necessary or desirable to conduct coordinate-type searches to identify those documents, persons, places, or things which meet a particular set of criteria. For example, management may have an urgent need for locating employees who can speak a certain language, have had certain types of experience, and are willing to travel. Conventional methods usually make it impractical, if not impossible, to conduct searches of this type.

Four general types of systems may be used for organizing information by conventional methods. The following is a description of each, together with an explanation of why each may sometimes fail.

1. Subject document files (fig. 3).

Definition: Documents arranged by subject categories, as in hierarchical subject classification

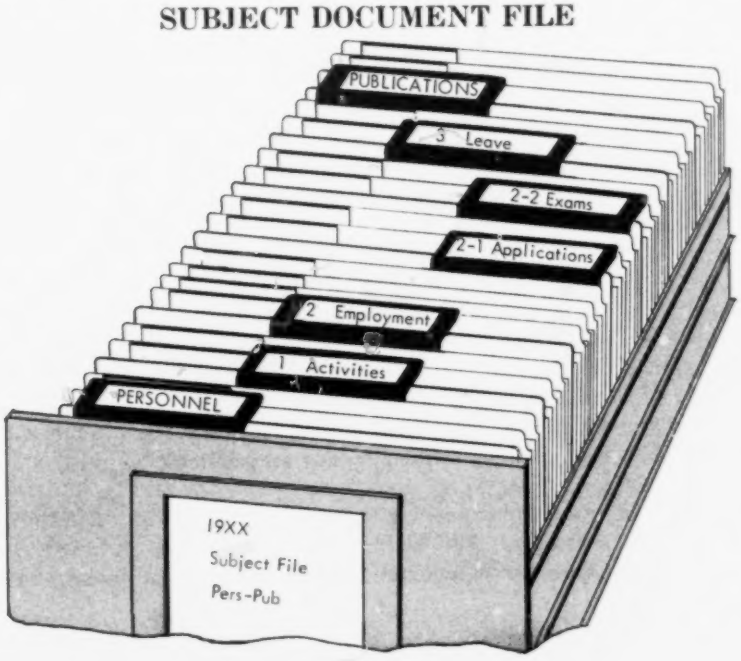


Figure 3

MANUAL SUBJECT INDEX CARD FILE

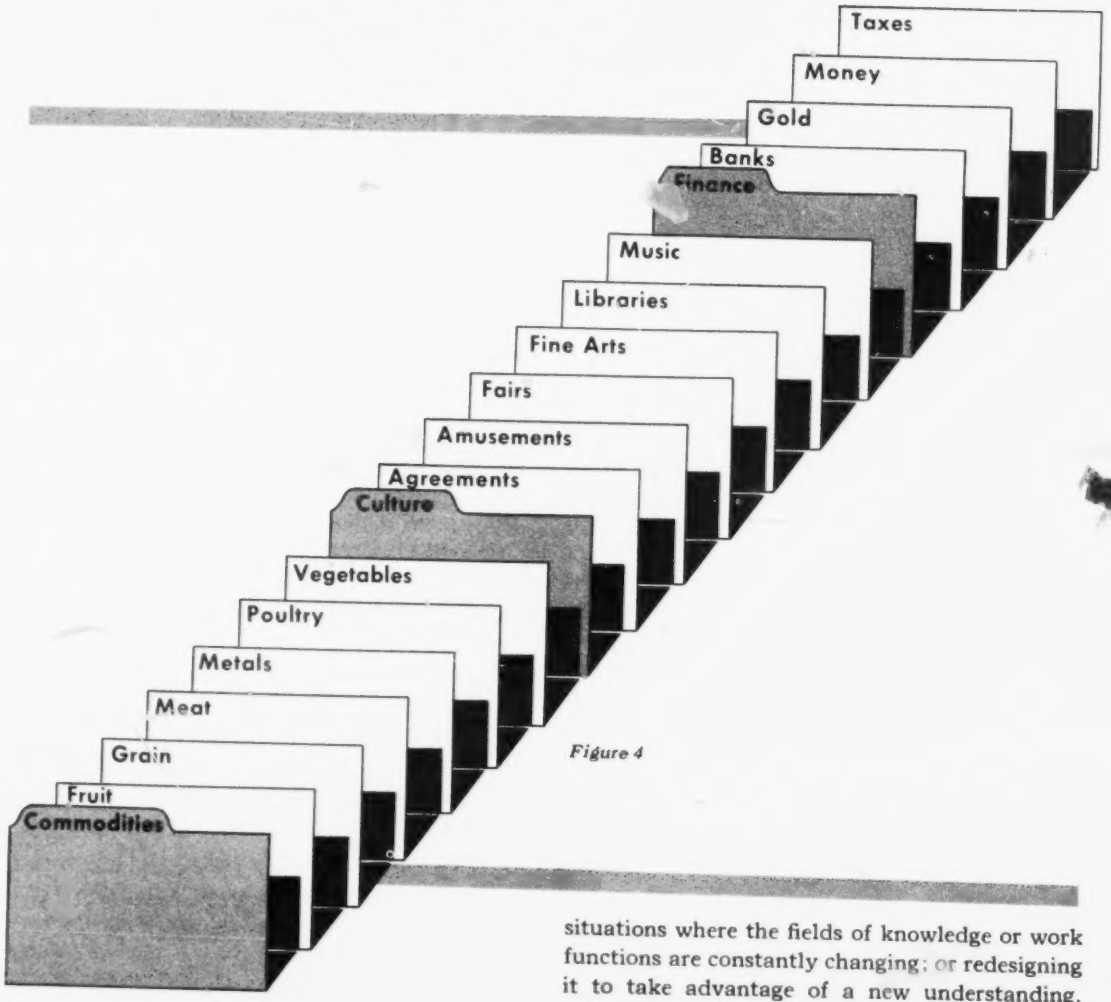


Figure 4

systems for correspondence folder files, library books, and other written material.

Significant problem: Developing a classification scheme that will satisfy the viewpoints, terminology, and needs of individual users in instances where the users have a wide variety of interests.
Why the system may fail: A hierarchical subject classification scheme needs to be directly related to the background and thinking processes of the users served. It is, therefore, virtually impossible to construct a classification scheme that will ideally serve the needs of a wide variety of interest groups.

Significant problem: Modifying the system in situations where the fields of knowledge or work functions are constantly changing; or redesigning it to take advantage of a new understanding, gained through additional experience with the system, of how the information should be organized.
Why the system may fail: Many times, the experience gained by using the system reveals shortcomings in the first arrangement that could be eliminated by reorganizing the classification structure. The rigid structure of a hierarchical classification scheme makes adjustments of this sort very difficult.

Significant problem: Classifying, filing, and retrieving individual documents in situations where they are often lengthy and involve numerous subject categories.
Why the system may fail: If an individual document relates to only one topic represented in the

subject filing system, there is no problem in filing it. But if the document has more than one subject, then cross-referencing becomes necessary. When such a situation is commonplace, the conventional system will tend to break down. A complex search involving several subjects can become a jungle of cross-references which makes the searching process very difficult, time consuming, and possibly unsuccessful.

2. Manual subject index card files (fig. 4).

Definition: Manual card files arranged by subject topics, as in a library's 3- by 5-inch subject heading card file.

Significant problem: Selecting subject terms that will be meaningful in the future.
Why the system may fail: Selecting subject topic terms that will always be meaningful and useful in the future is not only difficult but at times impossible. The problem is particularly thorny when conventional methods are employed.

Significant problem: Card preparation and updating costs.
Why the system may fail: Just the initial preparation and filing of manual index cards can be quite costly, especially if it is necessary to prepare and file several cards for each document; but to update a large file may be so costly that in actual practice it could not be done.

Significant problem: Detailed (deep) indexing of documents involving a large number of subject topics.
Why the system may fail: The physical limitations of index cards are a problem if a document must be indexed in depth. Detailed (deep) indexing of documents involving a large number of subject topics is difficult because of the size of the file that this practice would create. A card must be prepared for each subject in the document and a cross-reference prepared to all other related subjects. The structure of the card and the size of the file create barriers to fast and efficient searching. Collating these cards in a search is also very tedious and time consuming.

3. Case document files (fig. 5).

Definition: Documents arranged by case name or number, as in a personnel folder file.

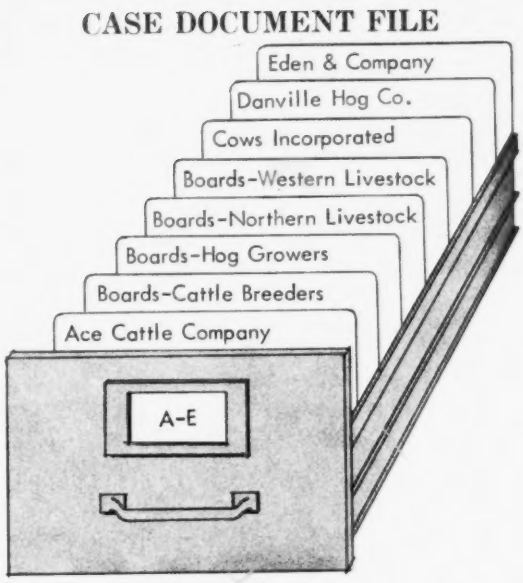


Figure 5

Significant problem: Searching large numbers of folders in situations where it is often necessary to correlate, compare, or analyze data, as in personnel selection and placement.
Why the system may fail: A case file containing large numbers of folders is very difficult to search if information must be correlated, compared, or analyzed. The physical problem of handling the folders prevents quick and easy reference. Every folder must be thoroughly analyzed from front to back before a complete job is done.

Significant problem: Locating or extracting specific items of data appearing at various places within the folder, in situations where the data is frequently needed for such purposes as answering inquiries and preparing reports.
Why the system may fail: The items of data in a document are usually not arranged for retrieval purposes but for easy preparation. When individual items must be located in a large number of case folders, the problem of pulling the folder and finding the item on the form becomes very tedious. A search of this type takes a lot of time and is subject to a large amount of human error in locating and transcribing information.

Significant problem: Locating precedent or policy material scattered among the case folders.
Why the system may fail: If material on precedent or policy matters must be located, usually it can be done only by making a search of the file

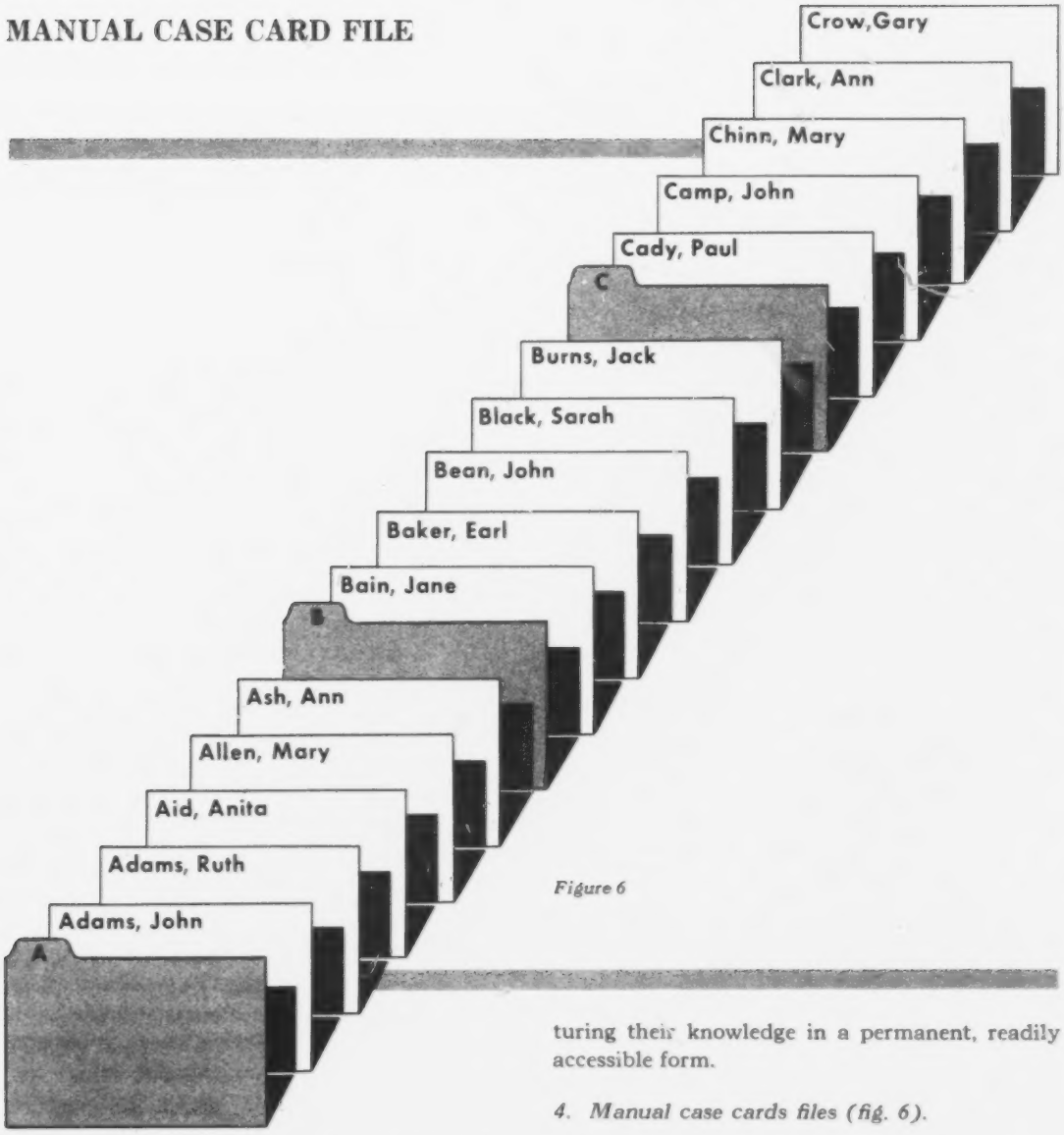


Figure 6

or calling upon the memories of employees who have had long experience in the subject matter field. Seldom is this type of information readily accessible in a separate section of the folder. The problems of interpreting precedent or policy matters are large enough but in addition, the typical case file has the disadvantage of requiring a tiresome, page-by-page, folder-by-folder search of the file for this type of information. Many organizations that depend upon the memories of long-time employees for such information are rightfully becoming, as these older employees retire, concerned with methods and techniques for capturing their knowledge in a permanent, readily accessible form.

4. Manual case cards files (fig. 6).

Definition: Manually prepared cards arranged by case names or numbers, as in a personnel data card file.

Significant problem: Cost of updating and preparing cards.

Why the system may fail: Card preparation and updating costs can be very high for such files. Each card must be manually prepared and individually inspected. As the size of the file grows, the point is reached where the cost of manually maintaining and updating the cards becomes exorbitant.

Significant problem: Losing or misfiling cards.
Why the system may fail: Manual case index card files must be properly designed and controlled to prevent loss of information. If the card for a certain item is lost, then the whole record of activity for that item is lost. Although methods of color coding, grooving, tabbing, and sequential numbering can make refiling so easy that even a newcomer to the system can do the job well, most systems are not this refined. Therefore, this possibility presents a severe limitation—particularly if the information is valuable.

Significant problem: Losing vital information through illegible hand postings and errors.
Why the system may fail: Whenever a file is manually maintained, a certain loss of information results no matter how many precautions are taken to prevent it. This is particularly significant in case card files because of the uniqueness of the information placed on each card. Preparing cards in this way makes verification for accuracy a very time-consuming and costly job. The best that can be hoped for is that most of the important mistakes are found and corrected.

Conventional systems always offer certain advantages, and if they will satisfy the needs of the users, they are often preferable to nonconventional systems. Chapter VI provides guidance on how to determine which of the two methods should be used. The following are the major general advantages of conventional systems:

- Usually simpler to design and operate.
- Require no special equipment.
- Permit direct access and often facilitate browsing.
- Input costs are usually lower.

Advantages and Limitations of Nonconventional Methods

A cost-benefit study should always be made before converting from a conventional to a nonconventional system. Nonconventional methods can, under the proper circumstances and application, result in one or more of the benefits described below.

Faster retrieval. This refers to the speed at which a user gets the exact information he needs to perform a task. Fast retrieval can be the significant element of a system when need is measured in seconds or minutes. For instance, if a child has swallowed poison and the antidote must be known immediately to save a life, speed is the most essential characteristic. Or if a policeman chasing a speeding automobile calls the station to identify the license number, again fast retrieval is essential.

Better information. This means information that is more complete, more accurate, and more current. For example, modern information retrieval systems can be designed that will reduce the chance that any pertinent information will be overlooked—a most important consideration in situations such as those facing the patent attorney or physician. Modern information retrieval systems make it practical to store and correlate more information and data since they usually have the capability to reduce masses of information to a manageable proportion more quickly than conventional systems.

Conserving users' time. How much time is spent searching for information through folders, reports, card files, book indexes, and other document files in an agency or field station? No one knows exactly, but in many situations it is far too much time. In some legal offices, for example, attorneys spend as much as 75 percent of their time searching for precedent decisions and the like. Modern information retrieval methods can save valuable users' time by reducing the man-hours spent in looking up, searching for, and correlating information needed to complete their tasks. Retrieval may be simple yet time-consuming, as in looking up individual social security numbers many times each day; or again it may be as complex and time consuming as in a one-time correlation of data to determine the possible cause of a missile failure.

Improve service. This refers to providing better agency service for the general public rather than to improving service within the agency for the direct users of the information

retrieval system. In some instances, it is possible to render service never before thought possible or to improve the service far beyond that which was possible when only conventional methods were available.

The full extent of the disadvantages and limitations of a nonconventional system may not become evident until the system has been in operation for some time. This is one of the reasons that a feasibility study is needed and that careful attention must be given all aspects of the system design. (For guidance in these matters, see chapter VII.) When compared with conventional systems, nonconventional systems generally have the following disadvantages:

- Require specially trained personnel to design and operate the system.
- Usually require special equipment.
- Often require use of special procedures and techniques to retrieve information.
- Input costs are usually higher.

Coordinate Indexing—Key to Many Nonconventional Systems

The concept of coordinate indexing—or concept coordination, correlative indexing, or multiple aspect indexing, as it is variously called—has been a major factor in removing the restraints imposed by earlier classification and indexing systems. All coordinate indexing systems have one feature in common: No attempt is made at time of input to limit the description of a document by classifying or indexing it under a major subject heading of two. Instead, large numbers of highly definitive indexing terms or data elements are employed, and the document is indexed under all entries that are pertinent. To retrieve information, the user selects those indexing terms or data elements that describe the items he is looking for, and the system quickly identifies all those that fit his description.

The key to the success of coordinate indexing is that all the descriptive information in the system is freely accessible, and no structuring of information takes place until a query is received. This permits an endless variety of on-demand searches to be made, each tailored to the precise interests and needs of the user.

Various types of equipment may be employed in coordinate indexing systems, as discussed in chapters IV and V; additional information on this subject is also included in chapters II and IV.

II. HOW COORDINATE INDEXING SYSTEMS WORK

Through the years, two traditional methods have been employed for organizing information by subject—hierarchical subject classification systems and manual subject indexing systems. The disadvantages and limitations of each were discussed in chapter I.

In hierarchical classification systems, the documents themselves are organized and arranged by primary subject categories and then further broken down by secondary categories, and so forth. Figure 7 illustrates two examples of hierarchical subject classification systems:

EXAMPLES OF HIERARCHICAL SUBJECT CLASSIFICATION SYSTEMS

Subject Numeric Filing System (office type)

ACCOUNTING
- 1 Accounts Current
- 2 Allotments
- 2–1 Symbols
- 2–2 Obligations
- 3 Disbursements
- 3–1 Loans

AUDIT
- 1 Assignments
- 2 Contract Audits

Dewey Decimal Classification System (library or office type)

600 APPLIED SCIENCE
- 610 Engineering
- 611 General Engineering
- 611.1 Equipment and Supplies
- 611.11 Tools
- 611.111 Cutting Tools
- 611.111.1 Stroke
- 611.111.11 Depth of Cut

700 ARTS AND RECREATION

Figure 7

The manual subject index file—such as the 3- by 5-inch card file found in most libraries—is often employed as a supplementary finding aid. Broad subject headings that are complete unto themselves are normally used, and the headings are arranged in alphabetical sequence. Typically, the card includes the title, date, author, and similar identifying information, perhaps plus a very brief description of the document. If the document is a book, usually several subject heading cards are prepared and filed alphabetically. Author and title cards may also be prepared and interfiled among the subject cards. The following are some examples of possible subject headings:

Automatic data processing
Correspondence management
Forms management
Information retrieval
Records retirement
Source data automation
Survey techniques
Work measurement

Principles of Coordinate Indexing

Coordinate indexing systems can be used to replace either or both of the hierarchical subject classification systems described above. The documents are identified and arranged by number, name, author, storage location address, or some other simple identifier. The index is usually a separate, highly manipulative, often mechanized file.

In a typical coordinate index, large numbers (sometimes thousands) of short terms are employed, most of which are not intended to be used alone but rather in any desired combination—"coordinated" to describe the various topics, concepts, aspects, characteristics, features, or attributes of the document or other item being indexed. These terms range from precise words and quantitative or qualitative data to abstract concepts or ideas. Both broad and narrow terms are used in the same system.

A zoologist, for example, might include such terms as those illustrated in figure 8 in his vocabulary of indexing terms:

SAMPLE VOCABULARY OF INDEXING TERMS

Africa
albatross
Antarctic
Arctic
Asia

bear
black
blue

capture
cat
color
conservation

deer
defense
diseases
dog
domestic
dorsum
duck/goose

eagle
ear
egg
elephant
Europe
exercise
exterior
eye

fish
food
fright

gestation
growth

habit
horse
housebreaking
hunting

instinct

leg
life span
lion

migration

1900 AD—present
1500–1900 AD
1000–1500 AD
Before 1000 AD

North America

obedience
offense

perception
population
preserve
price

reproduction
rescue
research
respiratory
rodent

shelter
size
South America
speed
strength

temperature—over 100°
temperature—80°–100°
temperature—60°–80°
temperature—32°–60°
temperature—under 32°

whale
white
worm

zebra
zooid

Figure 8

When indexing an individual document, all those terms that are pertinent are used to describe it. Thus, it can be seen that the description of the document consists of a group of interdependent terms that together comprise, in effect, a very brief abstract of the document.

In searching a coordinate index, one selects those indexing terms in the vocabulary that best describe the desired information. The index file is then searched to find any documents indexed under those terms.

Figure 9 illustrates the principles involved in searching a coordinate index. The cards represent indexing terms considered pertinent to a particular search question; the numbers on each card represent those documents indexed under that term. To find if there are any documents that would satisfy the search question, the searcher would look for particular document numbers that have been entered on all pertinent cards.

As in the indexing process, the searching process permits free coordination of a large number and wide variety of terms. For example, when desirable one can narrow the search by using more specific terms, or broaden the search base by dropping the more specific terms, or form new combinations of information or data by changing the configuration of the terms used in the search.

Types of Indexing Terms

Two types of indexing terms that may be used are as follows:

SEARCHING A COORDINATE INDEX

Figure 9

Keyword. The index terms consist of key words selected from the title or text of the documents. The indexing vocabulary is a by-product of the indexing process, and some form of control is usually exercised to keep the system manageable. The indexing of individual documents may be accomplished either by manual or machine (automatic) indexing methods.

Descriptor. A specially prepared vocabulary of indexing terms developed through a continuing process of analysis of the documents being indexed. The descriptors are usually formalized and controlled by means of a thesaurus. Indexing terms are manually assigned to individual documents from the approved list. Some of the terms selected to describe a particular document may coincide with keywords appearing within the document, while many will not.

Index File Arrangements

The index file is arranged in either of the two following ways:

By Document Numbers. A card or machine record is prepared for each document stored in the system, with all indexing terms describing the document recorded thereon. This is usually in coded form. Retrieval of information from the file involves sequential or serial searching, since the searcher must examine all the index records in the system to identify those documents that are assigned the terms used in the search.

By Indexing Terms. A card or machine record is established for each indexing term. When the indexer has decided which terms will be assigned to a particular document, the index records for those terms are selected and the document number is recorded thereon. Retrieval involves selective or parallel searching, since the searcher or the machine selects and examines only those records representing the terms used in the search.

Major Advantages of Coordinate Indexing Systems

- More Specific. Coordinate indexing makes it not only possible but practical to describe documents or other items in greater detail (depth) than conventional methods.
- More Adaptable. Coordinate indexes are far more adaptable to changing situations and unanticipated events than conventional methods.
- More Manipulative. Coordinate indexing makes it possible to quickly correlate and manipulate information and data in an endless variety of ways to achieve the desired search results.

Those desiring to install a coordinate indexing system have a wide variety of equipment choices. These include such manual types as the columnar, optical coincidence, and edge-notched card systems covered in chapter IV. Also, certain types of microform equipment, electrical accounting machine punched card systems, and electronic computers, described in chapters V and VI, may be used. For information about designing a coordinate indexing system, see chapter IX.

III. MICROFORM SYSTEMS

Microform is the general name for the various types and formats of microfilm and other media used for recording information in miniaturized form. In the past microform was used mainly for space-saving purposes; but numerous studies have shown that it is often less costly to place the records in the low-cost storage facilities provided by the Federal records centers. Today, however, microforms are assuming a new and far more important role in solving problems relating to information dissemination, storage, and retrieval.

How Microforms Help Solve Typical Information Problems

The following are some typical problems that can sometimes be solved or partly solved by the use of a microform. Moreover, it is not likely in any given situation that only one of these problems prevails, which largely explains the growing interest in microforms.

Problem: Document Accessibility

- Travel problem.
- Competition problem.

It is usually possible to keep near the users small collections of documents that occupy a file cabinet or bookcase. But the larger document collections, by necessity, are usually located at some distance from the users' area. This means that either the document or the user has to travel back and forth to the storage site.

Further, there are times when the same document is needed by more than one user, and each must wait his turn. These problems of course cause work delays. They also tend to reduce the usefulness of the information contained in the documents, since the users are inclined to try to do without unavailable documents if they can.

Both problems could be solved through the use of a microform system. Once the documents are converted to a microform, inexpensive duplicate sets could be placed in various locations in the users' work areas. A second choice, which solves the competition problem only, is to make film-to-film copies for multiple users who need to see the documents.

Problem: Document Servicing and Control

- Man-hour requirements for pulling folders and preparing document chargeouts.
- Man-hour requirements for filing returned documents.
- Man-hour requirements for following up on unreturned documents.
- Man-hour requirements for routine document maintenance.

If a microform system is used, inexpensive diazo copies of the documents can be made and given to the user instead of loaning the file copy. The user disposes of the duplicate copy when he is through with it. Thus there is no document chargeout and refile problem, and file maintenance is reduced to a minimum.

Because personnel costs are rising constantly and it is sometimes difficult to obtain file clerks, situations will be increasing where records managers must turn to microform to solve their problems.

Problem: Retrieval Speed and Costs

- Random lookup of individual items of data.
- Examination of graphic information.
- Scanning and retrieving information in textual documents and indexes.

In situations where a large volume of data can be readily converted to a microform, retrieval speeds sometimes can be increased for a very

small additional cost by use of this medium. This is particularly true of instances where retrieval involves random lookup of individual items of discrete data such as a social security number, date of birth, or street address.

If there is a continuing need for examination of graphic information—such as large maps, engineering drawings, or photographs—microform often will make the job faster as well as easier. Similarly, scanning or browsing through large collections of textual material and indexes is sometimes easier and faster if they are available in microform.

Overall retrieval speeds and costs often can be improved because a microform system makes it possible to store needed documents and data at the user's work station, rather than keeping them at a remote location.

Problem: Document Printing, Distribution, and Stocking

- High costs for printing, collating, and packaging of paper documents.
- Transportation and handling costs.
- Stock control and replenishment costs.
- Time-delay problem.

Many Government agencies discovered some years ago that the most economical and efficient way to reproduce, distribute, and fill individual requests for unpublished reports is by means of the microform. Federal agencies, within the Department of Defense in particular, are saving thousands of dollars each year by using the microform for reproduction and distribution of engineering drawings of military equipment.

Not only is it sometimes possible to reduce the initial printing costs, but significant savings can often be realized in handling and transporting documents. Stocking usually can be eliminated altogether, since the microform stored at the original source or at any distribution point can be used to reproduce on demand low-cost, film-to-film copies or enlarged paper copies. The original microform can be produced readily by photographing paper documents. However, with the advent of computer-assisted document preparation, editing, index preparation, formatting, and Computer-Output Microfilm (COM) equipment, direct publication of documents in microform is now possible. The computer output magnetic tape also can be used to automatically print paper copies. For many agencies, these new techniques offer the means for a substantial reduction in the time lag between document drafting and receipt by the users.

Problem: Computer Data Storage and Accessibility

- Storage and retrieval of machine language backup data.
- Storage and retrieval of static or semistatic data.
- Direct access to computer master file.

It doesn't take long for a computer to fill a reel of magnetic tape with data. If it is kept busy all day, the computer may have produced dozens of tape reels to add to the tape library. It is little wonder, then, that some computer installations have thousands of tape reels or millions of punched cards in their file and must often restrict the computer master files to summary data. While this backup data resulting from input processing and other machine runs is usually essential to system documentation, due to its great volume it is often too costly to retain the data in machine language and search it by computer. The Social Security Administration was among the first to use the microform and the first to procure a COM device to solve this problem.

While the computer provides the fastest and most accurate means for compiling, updating, and organizing static and semistatic data, the size and cost limitations of mass memories and time requirements often make it impractical to use the computer to retrieve data from these files. Often, the best current solution to the problem is to convert data recorded on magnetic tape to a microform by means of COM equipment. A special optical mark reader, called the "Foto" Optical Sensing Device for Input to Computer (FOSDIC), has been developed to read and process Hollerith-coded data on a microfilm copy of punched cards.

Such computer data bases as inventories, transportation schedules, rates, and special tables can be converted periodically to microfilm and then searched by means of standard microfilm readers. Where static information ties in with dynamic data maintained "on-line" with the computer, special remote terminals have been designed to permit the users to interrogate both data bases at the same time.

By necessity most large Automatic Data Processing (ADP) systems must use batch-processing techniques and access the master file on a cyclical basis—perhaps once or twice a day, once a week, or possibly less frequently. During the interim, the data is locked up in the tape reels and inquiries must wait until the next processing cycle comes around to be answered. By converting the data to a microform by means of COM equipment, inquiries and requests can be handled quickly and efficiently by nonskilled personnel equipped with microfilm readers.

Problem: Updating and Maintenance of Directives, Manuals, and Catalogs

- Total costs for individual updating of directives, manuals, and catalogs kept at numerous locations.
- Errors and delays in individual updating.
- Maintaining large, frequently used manuals and catalogs intact and in good condition.

The updating of maintenance and procedural manuals, catalogs, and similar publications can be a time-consuming and difficult problem if there are numerous publications and if they are maintained at numerous locations. Errors are made in entering the changes, while the insertion of some changes is delayed or never made at all. If the manuals and catalogs receive heavy use, as they often do in a maintenance shop, the pages are likely to be torn and lost. When detailed information is needed at the job site, the mechanic may have to copy the information by hand or remove a page.

In most agencies, no one knows exactly what this is costing or is aware of the full effects of not having current, accurate data on hand at each user location. However, in those instances where a detailed study was made, such as at some of the airlines, the savings were sufficient to pay for the cost of the microform system in a comparatively short time.

One of the ways to solve these problems through microform is to maintain a single master copy in cut-sheet form at a central point. Changes are entered in this master copy as they occur. The entire master copy is periodically rephotographed, reproduced in microform, and distributed to the users; whereupon, they simply dispose of the entire old copy. The microform readers are often equipped with a paper copier so that mechanics can make disposable copies to take back to their job sites when needed. In some situations the microform might also be produced through the use of the computer and COM equipment, as described earlier.

Problem: Procedural Bottlenecks

- Collection and transportation of large volumes of data.
- Verification of data on documents passing through the system.
- Logging documents.

Collection and transportation of large volumes of data such as questionnaires and reports can be a knotty problem if they are retained in their original paper form.

The U.S. Census Bureau, Department of Commerce, solves this problem by having the census questionnaires microfilmed at various locations in the field. The microfilm is then shipped to the headquarters office at Suitland, Md., where it is placed upon a FOSDIC microfilm optical mark reader. It converts the data to machine language code for processing by computers.

Several Government agencies receive large volumes of checks from the public. The checks can be microfilmed while being processed through the system in order to verify any data that may later be questioned. For similar reasons, organizations using Optical Character Recognition (OCR) equipment for computer input sometimes microfilm incoming documents.

The Division of Disbursement, Treasury Department, must maintain a record of each of the 1.5 million checks it issues each day. In the past, this was done by preparation of a paper record. Using COM equipment, the record is now produced directly from magnetic tape, making it possible to place the issue record for 102,000 or more checks on a single roll of microfilm. Duplicate microfilm copies of each month's veterans' benefit check issues are sent to Veterans Administration regional offices throughout the United States where the microfilm is used to answer thousands of inquiries a month, conduct postaudit operations, obtain a historical record of payments in specific cases, and locate addresses.

If it is necessary to log incoming and outgoing documents, microfilming is usually a much simpler and cheaper method than keeping records by hand. Many libraries use this technique for charging out books. Equipment manufacturers have developed lightweight portable cameras, including some that are battery operated, that add to the practicability of using a microform.

Problem: Storage and Handling of Large and Nonstandard-Size Documents

- Special equipment needs.
- Folding and unfolding of oversize documents.
- Storage of documents with irregular sizes and shapes.

Oversize documents such as tracings, drawings, and maps can be recorded on microfilm to eliminate the problems of special equipment requirements and the need for unfolding and folding the documents each time they are used. However, the original documents must conform to certain quality standards in order to produce a satisfactory microfilm substitute.

Documents having irregular sizes and shapes can be reduced to a uniform size through microfilm. Improved color microfilm is available if color is a significant factor.

Problem: File Integrity

- Errors in filing.
- Accidental loss or destruction of information.
- Alteration and obliteration of information.
- Users' failure to return documents.

Errors in filing occur in spite of the best efforts of file supervisors. If the file is a large one, it may be days, months, or years before a missing document turns up. Whenever a document is removed from a file and forwarded to a user, it might be lost in transit, accidentally destroyed, damaged, or not returned. These, of course, are serious risks when dealing with important documents such as those affecting individual rights and claims.

Often the best way to insure absolute file integrity is to convert documents to a microform system. The user is provided access by furnishing a film-to-film copy or an enlarged paper copy for his use.

Problem: Document Acquisition

- Rising cost of hard copy publications.
- Acquisition of rare or unique documents.

The rising costs of publications printed in paper copy are making it necessary for many libraries, offices, and others to curb their document-acquisition programs. In those instances where a document is available in either paper copy or microform, savings of 70 percent or more can usually be realized by purchasing microform.

There are also times when desired documents are out of print. If such documents are needed urgently, the simplest and generally cheapest way is to make microform copies.

Problem: Document Preservation and Protection

- Prevention of wear and defacement of valuable, irreplaceable documents.
- Protection of indispensable operating records against a disaster.
- Protection of classified documents.

The Library of Congress and the National Archives use microfilm extensively for preservation of important documents. The microfilm copies are made available to scholars and researchers, not the original documents.

Microfilm is used by many agencies for protection of indispensable operating records against a fire or national disaster. The film is usually kept in a remote, protected depository that in most instances is equipped with machines and supplies for making film-to-film copies or paper enlargements. The original copies of classified documents may be microfilmed so that either the original or copy of the document is always secure.

Problem: Equipment and Space for Document Storage

- Availability of adequate space to house documents.
- Space costs.
- Equipment costs.

While space and equipment savings are often an important factor in a microform cost-benefit analysis, microfilming can seldom be justified for this purpose alone.

Prerequisites for a Successful Microform System

For a microform to serve as a satisfactory substitute for paper copy, it must be as legible and easy to use as its paper counterpart. Microform system success depends upon such factors as condition of the original documents, the film, the camera, the camera operator's work, the quality of film processing, the suitability of the microform type, proper storage and handling of the microform, the adequacy of viewing equipment, and the ability to quickly locate information within the microform record. A weakness in any of these areas may cause the system to fail.

The single most critical factor is the condition of the document. Not only does this largely govern the quality of the finished microform, but it is a major cost factor in the filming operation. Typical problems are poor contrast between the reading matter and the paper; extremely fine lines or printing; lack of uniformity in color, size, and thickness of documents: intermingling of one-sided and two-sided documents; the need for removal of staples, pins, and other fasteners; and the need for sequence checking and screening to remove extraneous material.

Within the next 10 years it can be expected that many of the existing large-folder file systems in the Federal Government will be converted to microform. Steps should be taken as soon as possible, therefore, to clean up and revise such systems so that the essential papers will be susceptible to low-cost, high-quality microfilming. Careful attention should also be given to the planning and maintaining of any new, long-term paper document files so that they too may be readily converted to a microform should this later become desirable.

Types of Microfilm and Cameras

Normally, the initial step in any microform system is the recording of document images on roll microfilm having a silver base. This master film, in which images appear in a negative mode, is then used to produce duplicate reference copies as needed. The copies may also be silver films, but if widespread duplication is necessary the lower cost ammonia-developed diazo films are commonly used. A third type, thermally developed vesicular films, may also be used for producing reference copies.

While the original microfilm master is normally in roll form of 16 mm, 35 mm, 70 mm, or 105 mm width, the reference copies are often cut into small pieces for use in systems employing unitized microform media. These include strips, chips, microfiche, microfilm jackets, and aperture cards, which are described later in this chapter.

Four main types of cameras are used in the original filming operation. See figure 10. These are as follows:

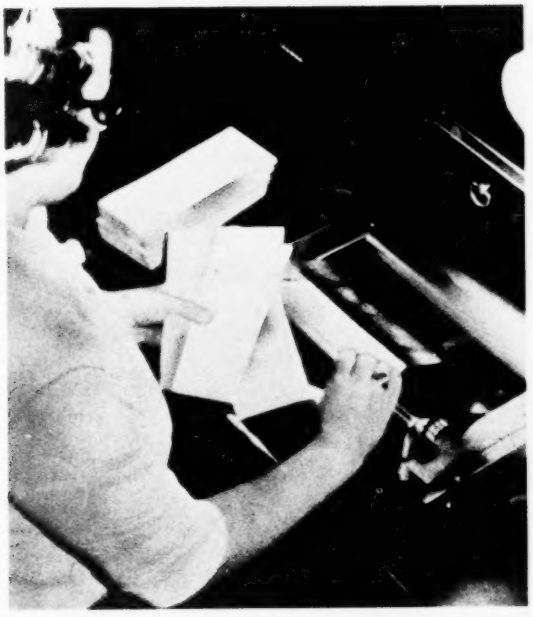
Planetary cameras are employed for obtaining high quality microfilm of engineering drawings, maps, and assorted other documents that cannot be satisfactorily filmed by a rotary camera.

Step-and-repeat cameras are used for direct film-

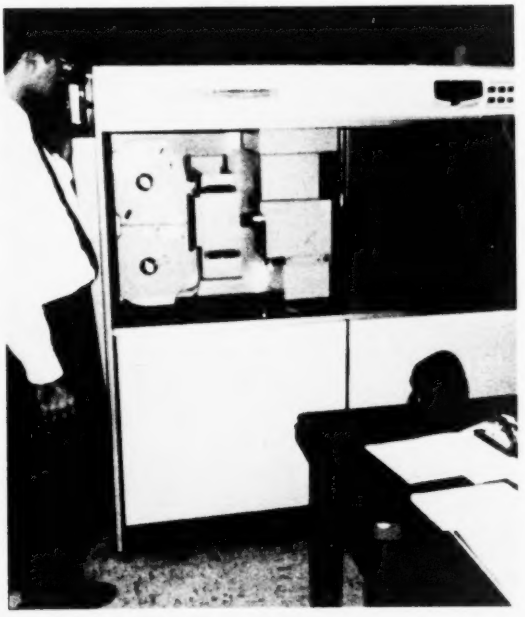
TYPES OF MICROFILM CAMERAS

Planetary Camera

Step-and-Repeat Camera



Rotary Camera

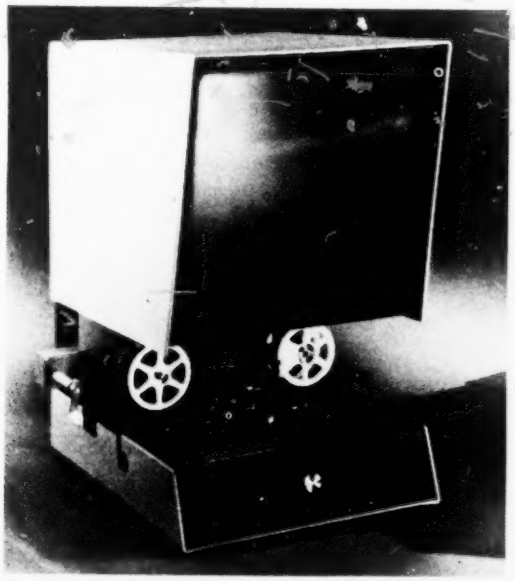


COM (Computer-Output Microfilm Device)

Figure 10

HAND-DRIVEN ROLL MICROFILM READERS

Washington Scientific Industries
Model RH Portable Reader



University Microfilms Reader Model 1414

The University Microfilms
Model 1212 Reader

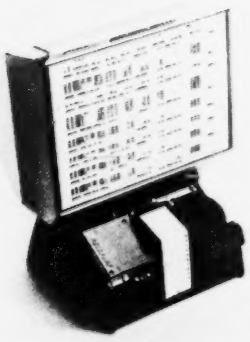
DASA Corporation's Mark I Model U Reader

Figure 11

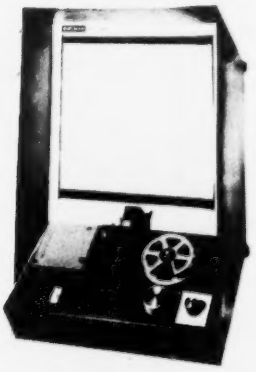
MOTORIZED READERS FOR ROLL MICROFILM

The Information Handling Services Satellite IIW Reader

The Recordak Motormatic Reader, Model MPG



The Ednalite 1624 COM Reader



The Dietzgen 4317 Reader

Figure 12

ing of documents in the multiple-row microfiche grid format. (Microfiche may also be constructed by cutting 16 mm or 35 mm film into strips and placing the strips in microfilm jackets or arranging them in rows on a special frame or sheet of clear film.)

Rotary cameras are used for filming printed and other documents of uniform size and color where ordinary film quality will suffice. They are largely automatic, thus permitting higher input speeds and use of unskilled operators.

Computer-Output Microfilm (COM) devices record computer-produced data directly onto microfilm, thereby bypassing the preparation of paper documents altogether. These devices can also add automatically to the microfilm copy the bars or code lines, image marks (blips), or photo-optical binary codes often employed to assist in the retrieval of documents or data.

Factors Affecting the Choice of the Type of Microform System

The choice of which microform system to select is governed by many factors. Mainly, these are the height and width of the documents, the number of pages per document, the total volume of documents or data, organization of the file, nature and extent of changes and additions to the file, number and location of the users, nature of

FILM CARTRIDGES AND CASSETTES FOR MOTORIZED READERS

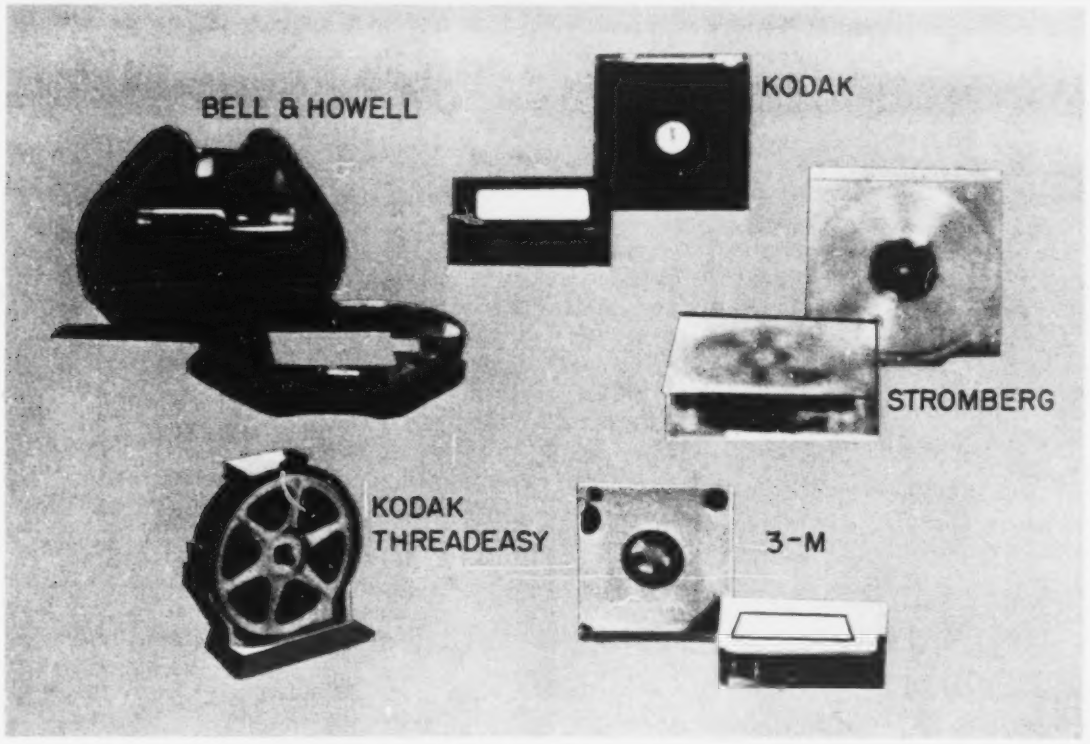


Figure 13

the reference activity, reference rate, retrieval speed requirements, and requirements for producing film or enlarged paper copies. Information on gathering the necessary data for system requirements, analyzing user needs, and selecting the right method and equipment class is provided in chapter VII.

Guidance on selection of particular manufacturer's equipment is contained in the records management handbook, *Microform Retrieval Equipment Guide*. A description of a number of systems employing microforms is included in the records management handbook, *Information Retrieval Systems*.

Types of Microform Systems

The following are descriptions of the various types of microform systems, together with a brief summary of the main advantages and limitations of each. Most of the microform readers mentioned are also available in reader-printer models that can produce full-size paper copies of the documents.

Conventional Roll Microfilm. These include systems using hand-driven microfilm readers and standard microfilm reels, as illustrated in figure 11. Flashcards or flash targets are used to separate file segments or pages. (Figure 14 depicts a sample of a flashcard used on roll microfilm.) Conventional roll microfilm systems are well suited to storage or protection of documents for archival, administrative, legal, or security purposes, and other situations where there is a very low reference activity. The main limitations of conventional roll microfilm systems are slow retrieval speeds and inconvenience to the user. The microfilm must be hand threaded through the reader, a slow and tedious operation. The user must then hand crank the film and scan the reader screen image by image until he finds the desired document.

Motorized Roll Microfilm with Mechanized Image Locator Aids. In both this and the system that follows, most of the microfilm reading equipment has been improved in three ways. First, a motor usually with both high and low speeds has been added; second, film cartridges or cassettes have been substituted for standard microfilm reels, and the reader has been made self-threading; and third, new techniques or devices have been employed to aid in locating desired film images.

Except for conventional roll microfilm systems, the motorized roll microfilm systems with mechanized image locator aids are generally the lowest in overall costs. They offer particular advantages for lengthy documents or record series. They can be successfully employed for the reproduction, dissemination, storage, and retrieval of catalogs, manuals, and publications, in which event many of the advantages described below for microfiche apply.

Figure 12 shows some typical motorized (mechanized) roll microfilm readers and reader printers while figure 13 provides examples of the various types of cartridges or cassettes employed. The mechanized image locator aids are of three types, as follows:

- Bars or code lines superimposed between images on the film that, when matched with a corresponding scale on the reader screen, can usually localize the search to within ten images or less, in a sequentially arranged numerical or alphabetical file.

- Film pull-down (linear location) aids that employ microfilm readers incorporating an odometer-like device for finding images on the basis of their linear location on the film. As in the system using image counting, this one depends upon the user's knowing or separately looking up the location of the desired image.

- Image count aids, which consist of marks (blips) superimposed beside each film image for use on a reader that has a photoelectric counting device. To locate an image, the user must know or separately look up the image location number for the document he desires to view. He enters this number on the reader keyboard, and the film automatically moves through the reader and stops when it reaches that number.

Figure 14 depicts examples of roll microfilm employing these three mechanized locator aids.

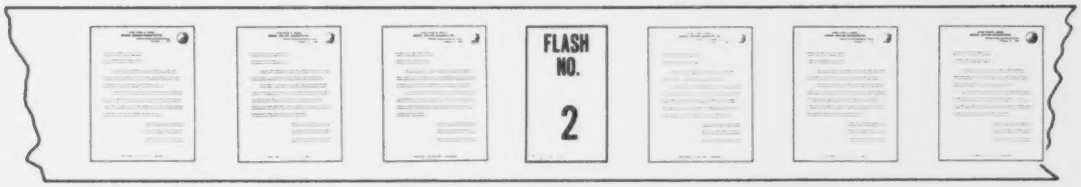
The use of the cartridges and cassettes with self-threading motorized microfilm readers has substantially improved the ease and convenience in the use of roll microfilm. The image-finding aids are a real boon to retrieval speeds in situations where they can be satisfactorily applied. Of the three techniques, the film pull-down (linear location) is usually the least costly and can be incorporated into a system quite easily. The bar or code systems are the next least costly and somewhat more difficult to incorporate into a system.

All three image-finding techniques have certain limitations. Bar or code line systems can be used only where the file is sequentially arranged by numerical or alphabetical identifiers and the user is conducting his search on the same basis. While the film pull-down (linear location) and image count techniques permit the documents to be in random sequence, a separately maintained list or index may be required for use in determining the proper microfilm roll and image location. Systems employing the image count technique require microfilm readers that are more complex and hence normally more costly than those used in the other two.

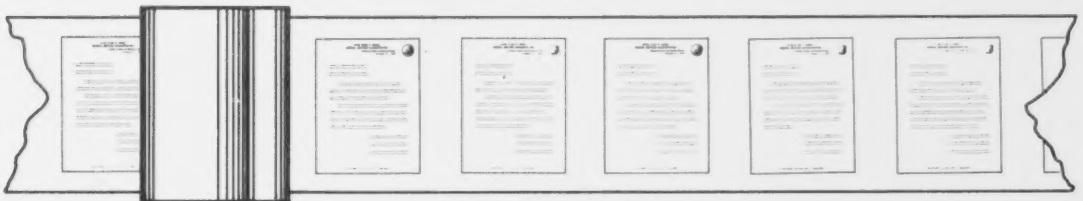
Special Note on Changing or Adding to Roll Microfilm

Most roll microfilm systems have one problem in common—changing or adding to previously filmed records. There are three methods for doing this, and none may prove entirely satisfactory. However, under certain circumstances, one or more might prove practical. The first and least likely method (except for publication of catalogs, manuals, listings, and COM produced items) is to retain the original documents, make the changes, and periodically refilm the entire file. A second but not always practical choice is to film the changes or additions and splice the new film onto the old film. A third method is to film the changes

INDEX METHODS USED IN 16mm FILM



Flash Card



Film Pull-down

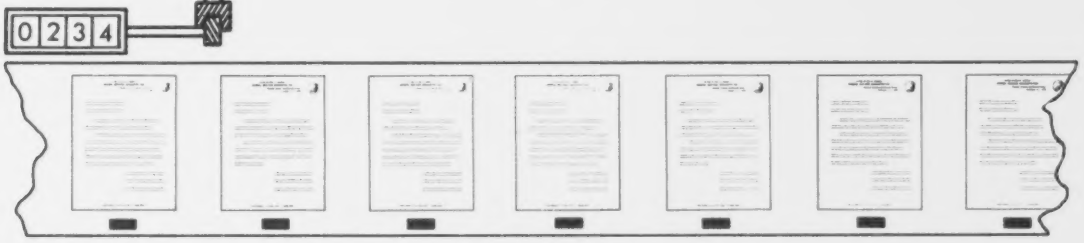
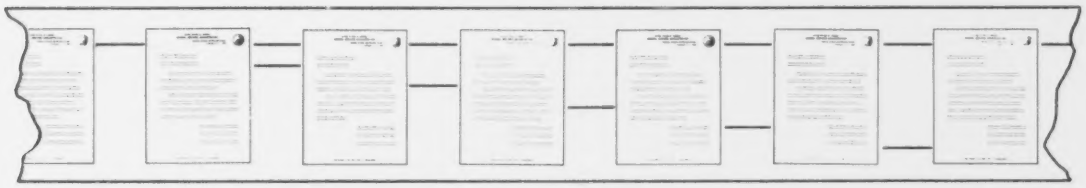


Image Count



Bar or Code Line

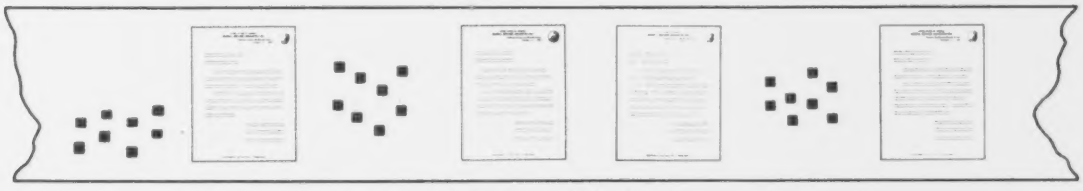


Photo-optical Code

Figure 14

and additions as they are received, add the new rolls to the microfilm collection, and maintain a separate index or locator record (preferably computer maintained and produced) showing the location (microfilm roll and possibly image number) of both the old and new images. This requires the user to make a double lookup, but this may prove to be only a minor handicap.

Roll Microfilm with Photo-optical Binary Code. This type of coding system can be used to conduct computer-like searches. Figure 14 depicts a sample of photo-optically coded roll microfilm. Such document descriptive data as titles, names, dates, numbers, and subject topics can be recorded in photo-optical binary code format on the film, thus permitting the user to automatically conduct both simple and complex or coordinate-type searches. Depending upon the features of the particular equipment, search entry is made through a keyboard, dials, or a machine record such as edge-punched cards.

The major advantage of the motorized roll microfilm system with photo-optical binary code is that it permits the user, while conducting the computer-like search, to simultaneously see the documents involved.

The major disadvantage of these systems is the cost. Except where COM equipment is employed for preparing the microfilm, the input costs are usually greater. The retrieval equipment costs more than that used in most other microfilm systems and is somewhat more difficult to operate. Unlike computer systems, the binary optical code, once recorded on the film, cannot be changed. Further, unless the file can be broken down into separate autonomous groups and the individual searches confined to a single group, the time required to conduct individual searches will increase as the file grows. This could result in a need for additional equipment and personnel, and thereby tend to offset the initial advantages of the system.

Microfilm Strip Systems. Microfilm strip systems employ roll microfilm cut into segments for storage of multipage documents. Three general manual methods used for storage and retrieval of the strips are: (1) maintenance in separate small metal or plastic containers; (2) attachment of the strip to the edge of a card or sheet containing full-size written information; and (3) attachment of the strips to plastic sticks about a foot long maintained in horizontal racks for rapid removal and refiling. The first two have received limited use for dissemination, storage, and on-demand reproduction of lengthy documents, while the third has been used primarily for storing and retrieving information and data contained in such listings as a directory or catalog. Figure 15 depicts a microfilm strip attached to a plastic stick, and the special storage rack and reader used for this type of strip system.

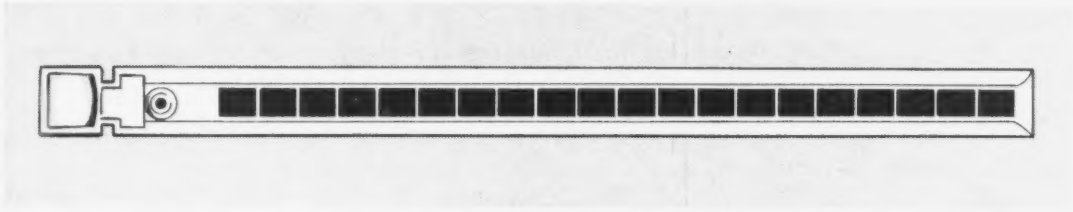
All three techniques provide a means for unitizing microfilm so that the individual documents or parts thereof may be independently selected, viewed or copied, and refiled. The third technique facilitates storage and retrieval of lengthy listings by making it possible to keep them in a very small space while at the same time permitting random, fast access to the information. However, an actual test is always needed to determine comparative retrieval speeds.

The major problem with the first type is that of physically handling the strips—opening the container, hand threading it through a reader or splicing it onto another length of film, and returning it to storage. The main problem with the second type is that it, too, is somewhat awkward to handle and can only be used in certain microfilm readers. The main limitation of the third type is the cost of preparing and mounting the film and purchasing the special reader required to view the film.

Microfilm Chip Automated Systems. These systems, as illustrated in figure 16, usually employ small pieces of cut microfilm that are often stored in cartridges or cells and manipulated by means of electronic circuitry and electromechanical devices. A keyboard or other device is used to conduct searches. These systems have been used primarily to meet the need for high-speed retrieval of short documents (one to three pages, generally) from extremely large files.

In some systems, a considerable amount of photo-optical binary coded data can be entered on the chip, while in others only a document number or address can be recorded. In one system

EXAMPLE OF A MICROFILM STRIP SYSTEM



The Microfilm Stick

Special Storage Rack and Reader

Figure 15

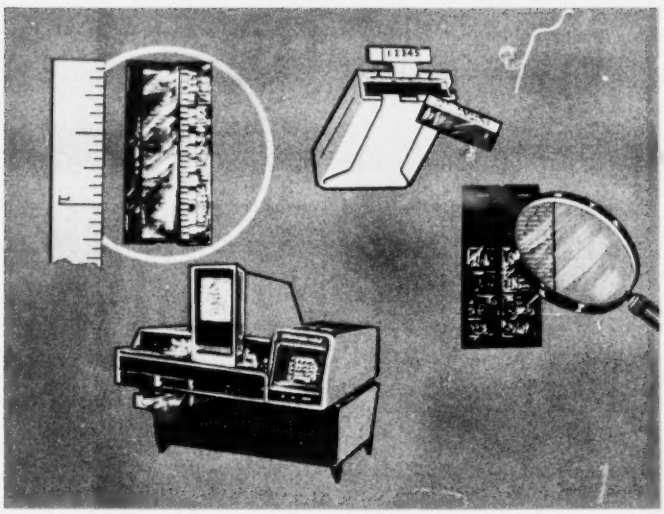
there is an iron oxide coated strip for recording data by means of a magnetic binary code, as on the magnetic tape used on computers. Microfilm chip systems are quite complex, usually involving rather high equipment costs, and thus have not been used as extensively as some of the other systems. The hardware is generally not available off the shelf but must be custom engineered.

Microfiche. Microfiche, as illustrated in figure 17, are sheets of microfilm containing multiple rows of micro-images arranged in a grid pattern. Microfiche are particularly well suited to the reproduction, dissemination, storage, and retrieval of documents or records having a total length of 20–98 pages or having chapters, sections, or parts of that length; they can also be used for longer documents, of course. Microfiche are sometimes used for storage of case-type material, such as hospital records.

The two most commonly used microfiche formats are both about 4- by 6-inches in size. The formats shown in figure 17 (60 pages per microfiche) was adopted in 1965 as the Government standard for reproduction of scientific and tech-

AUTOMATED MICROFILM CHIP SYSTEM

Figure 16



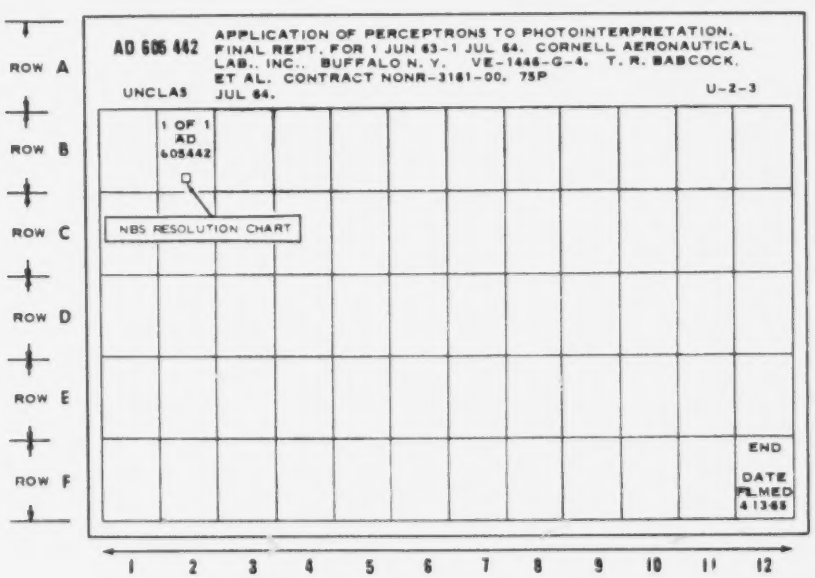
nical documents. Another format (98 pages per microfiche) has recently been increasingly used by both industry and Government. Figure 18 describes some of the wide variety of microfiche formats and reproduction ratios in use today, including high reduction (HR) ratios.

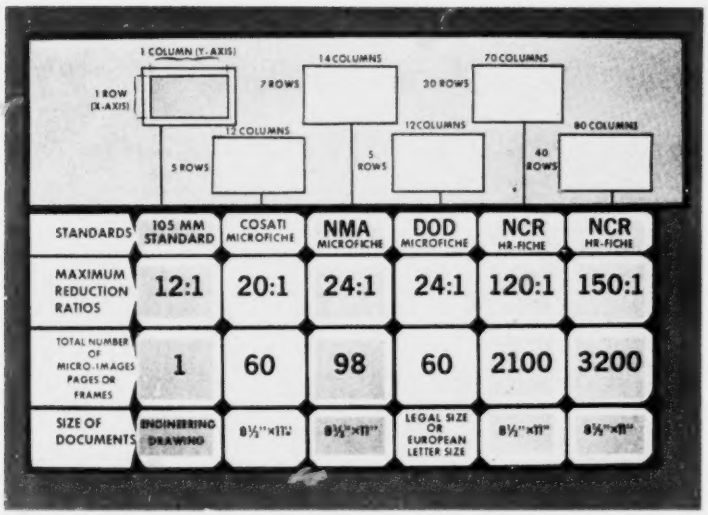
One of the major advantages of the microfiche is a possible savings of 70 percent or more to the user in acquisition costs in instances where a document is available in both microfiche and paper form. Another advantage is the elimination of document warehousing problems, since low-cost copies of microfiche can be produced at any point on demand. In many situations the most significant advantage is the savings in time and costs for packaging, shipping, storing, and retrieving documents.

Probably the major disadvantage of the microfiche is the relatively high input cost, which may make this type of microform uneconomical for internal application within a single office. However, if the documents are widely distributed, input costs can become quite insignificant. Another disadvantage is that there has been no prac-

SAMPLE OF A MICROFICHE

Figure 17





STANDARDS	105 MM STANDARD	COSATI MICROFICHE	NMA MICROFICHE	DOD MICROFICHE	NCR HR-FICHE	NCR HR-FICHE
MAXIMUM REDUCTION RATIOS	12:1	20:1	24:1	24:1	120:1	150:1
TOTAL NUMBER OF MICRO-IMAGES PAGES OR FRAMES	1	60	98	60	2100	3200
SIZE OF DOCUMENTS	ENGINEERING DRAWING	8½"×11"	8½"×11"	LEGAL SIZE OR EUROPEAN LETTER SIZE	8½"×11"	8½"×11"

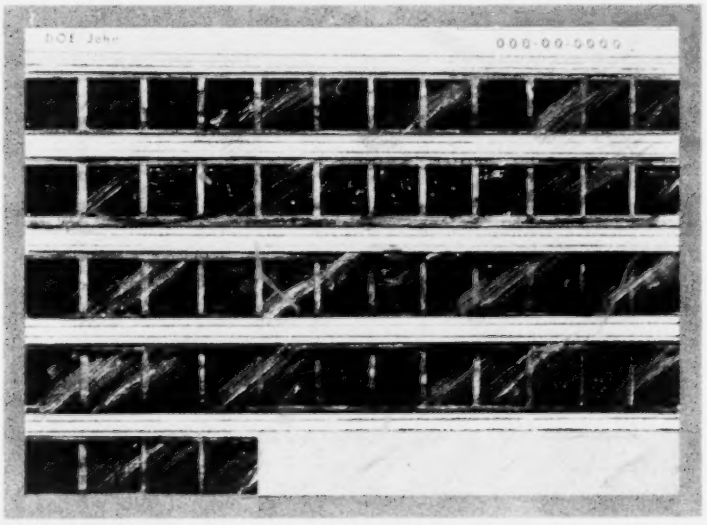
COMMON MICROFICHE FORMATS AND REDUCTION RATIOS

Figure 18

tical, inexpensive method for changing or adding to individual microfiche (up to the time this handbook was prepared). If updating is required, the alternatives are similar to those described above under the heading, "Special Note on Changing or Adding to Roll Microfilm."

Still another factor limiting the use of the microfiche is that special readers are required at every point of use; and even though inexpensive readers are available, the overall equipment investment may be substantial. However, as the use of the microfiche is extended to more and more document series, the readers may eventually become standard office equipment. Another possible disadvantage is that some users feel that further improvements are needed in the readers in order to make the viewing more convenient and comfortable.

Microfilm Jackets. Microfilm jackets are transparent carriers with one or more sleeves or pockets for holding strips of microfilm, as shown in figure 19. The entire jacket, with the microfilm inside, is placed in a reader for viewing. Film-to-film copies and paper enlargements may be made without removing the film from the jacket. To get the best results it is necessary to use one of the newer "thin film" jackets.



SAMPLE OF A MICROFILM JACKET

Figure 19

SAMPLE OF AN APERTURE CARD

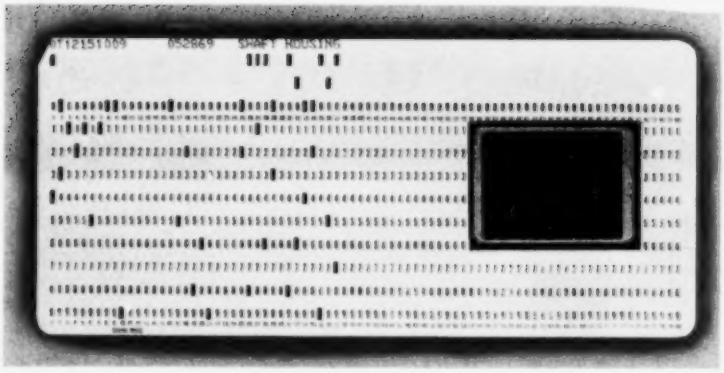


Figure 20

The major advantage of the microfilm jacket is that new images may be added, thus making it particularly suitable for active case-type records. It is compatible with the microfiche and can be used in the same types of readers and film-to-film copiers, and thus has many of the advantages noted above for the microfiche.

The major disadvantage of the film jacket is the time required for inserting individual micro-images into the sleeves of the jacket; however, special equipment has been developed for this purpose to make the task much easier.

Aperture Cards (Microfilm Electric Accounting Machine Punched Card). These cards, illustrated in figure 20, are standard punched cards (or edge-notched cards) with windows containing micro-images. The window is usually designed to accommodate one large document, such as an engineering drawing, or as many as eight or 10 letter-size pages, which in the case of the punched card, would require 22 card columns of space. This leaves over 50 columns for recording data such as the document number, description, and date in machine-coded form. There are also aperture cards containing sleeves as in microfilm jackets for inserting and adding images.

One of the major advantages of the aperture card is the convenience in filing, retrieving, and adding to the file. Another advantage of aperture card systems is the capability for using mechanical devices for sorting and selecting individual cards, while at the same time permitting manual filing and selection of cards. Still another important advantage is the savings in time and cost for duplicating, shipping, handling, storing, and retrieving documents. Further, there is available a

SUPERMINIATURE (HIGH REDUCTION) MICROFORMS

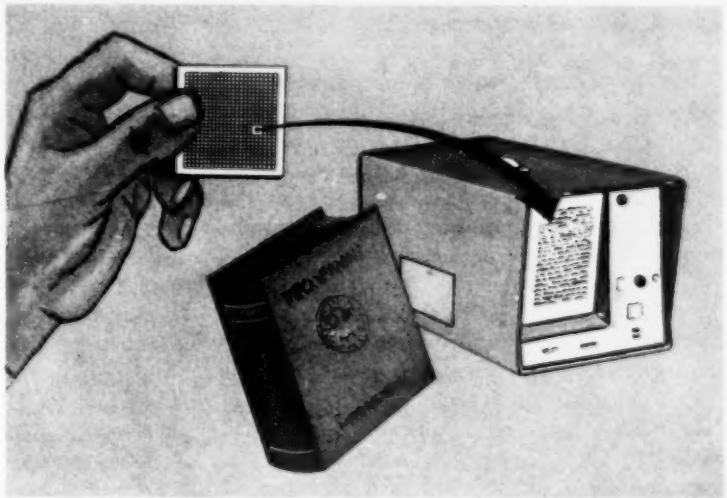


Figure 21

VIDEO RECORDING SYSTEM

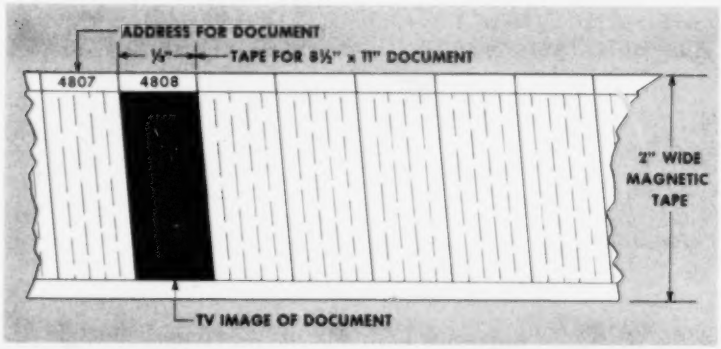
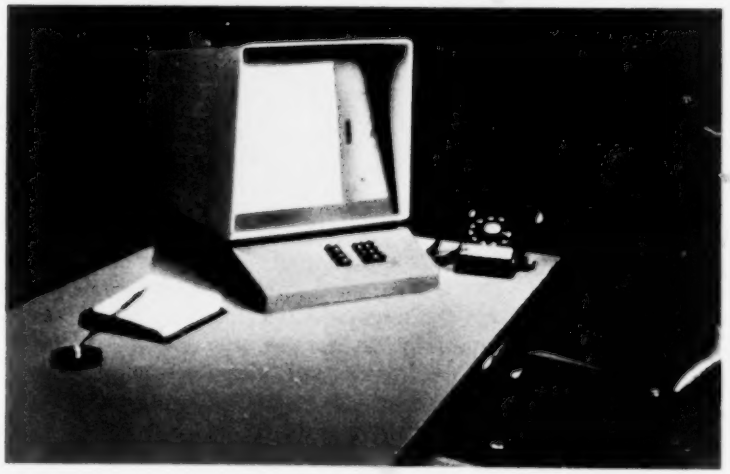


Image Storage



Desktop Viewing Equipment

Figure 22

wide variety of equipment to satisfy the needs of the smallest to the largest user.

The major disadvantage of the aperture card system is the relatively high input cost involved in the filming, keypunching (or edge-notching) of the cards, and mounting of the micro-images in the apertures. Therefore, as in the case of microfiche, the cost may make such systems uneconomical for internal application within a single office. Further, extensive machine sorting and selection of the aperture cards may not be practical if the file is a very large one. When punched card equipment is used for card sorting and selecting, it is usually modified in order to minimize damage to the micro-images; or a duplicate "slave" deck, which does not contain the micro-images, is created for use in the punched card machines.

Superminiature (high reduction) Microforms. Superminiature microforms and those referred to as ultraminiature microforms (ultrafiche) employ a reduction ratio much higher than those used for ordinary microforms. (See figure 21 for an example of a book of more than 1,000 pages reduced to one ultrafiche, and the special reader required for viewing the images.) The standard reduction ratios in use today readily permit the recording of 2,000 to 2,500 letter-size pages on a 100-foot roll of microfilm (and in some systems, up to 4,000 pages per 100-foot roll). Reduction ratios as low as 10 to 1 (10X) are used for newspapers and as high as 42 to 1 (42X) are used for COM produced listings and cancelled checks. Superminiature microfilm, on the other hand, employs reduction ratios of approximately 200 to 1 (200X) and higher.

The major advantage of superminiature microforms is the further savings in space and shipping costs resulting from the greater compactness of the micro-images. Superminiature microforms make it possible to store an extremely large collection of documents close to the users, or possibly within the viewing equipment itself.

The major disadvantage of the superminiature microforms is the initial cost of preparing the master copy. However, as in the case of the microfiche and the aperture card, this cost may not prove excessive if there are a large number of users at various locations who use the same information over and over again. Another disadvantage is the lack of compatibility between this and any other microform media. Special readers with optics compatible with the very high photographic reductions are required.

Video Recording Systems. These systems employ the basic techniques and equipment used in recording television programs, as illustrated in figure 22. Documents are placed under a camera and magnetically recorded on video tape or other media. There is a separate track for recording the document's number or other identifier. Retrieval is accomplished through a keyboard or by preparing a machine record such as a punched card that is fed into the retrieval device. Images of the retrieval documents may be viewed on remote terminal cathode ray tube (CRT) screens, or enlarged paper copies can be produced.

The major advantages of the video recording systems are the instant recording and inspection of document images; the ability to add or delete documents; the ease of use; and the relatively fast retrieval speeds. Video recording systems have not been in use long enough to fully evaluate their performance and potential. However, the major disadvantages appear to be the relatively high systems cost; the need for special skills in planning, operating, and maintaining the system; and the need for special work procedures and routines to compensate for the lack of a practical means for gaining random access to the file.

Special note on mechanized devices (miscellaneous card selectors) for storage and selection of microfiche, microfilm jackets, aperture cards, and other unit records.

There are numerous electromechanical devices that permit selection of individual unitized microforms by means of a keyboard. The smaller ones have trays holding approximately 1,000 items each, which can be interconnected and operated through a single keyboard. Typically, the individual items are notched along the bottom edge, and the selected item pops up when its identifying number or location address is entered on the keyboard.

There are also very large units, some of which can be accessed through remote terminals equipped with keyboards and CRT displays. Some also have the ability to perform limited coordinate-type searches.

The major advantages of these devices are that they reduce physical strain, eliminate the need for interfiling as microforms are returned to the file, and make possible an increase in retrieval speed.

The major disadvantage is cost. To justify the purchase of such equipment the file must be very active, but not more so than one person per keyboard could handle. Thus, the limited access to the file could pose a serious problem in times of peakloads, expanded reference activity, or machine breakdown.

Microform—Computer Combinations

The motorized roll microfilm systems with photo-optical binary code and the microfilm chip systems combine in a single medium both machine-readable data and document images for simultaneous searching and viewing of the micro-images. Further, it is possible to use any of the various types of microform methods and equipment described earlier in combination with a computer. There are, however, an increasing number of microform devices specially designed for direct use with the computer.

Computers, as explained in chapter V, can perform complex coordinate and other types of logical searches, as well as other forms of data manipulation, at fantastically high speeds. However, storage of very large volumes of data on-line can be extremely expensive; and since computers can only work with information that has been converted to a machine language code format,

their capability for storage and presentation of graphics and large masses of data is rather limited. The situation is much the reverse for microforms, of course. Consequently, the computer and the microform can often be used to complement each other very effectively by maintaining the low-volume index data (or dynamic data) on-line with the computer and the large volume of information and graphics (or static data) in microform. Finally, a communications link—either human or part human and part machine—or all machine—is provided to permit the two to work as a team.

Roll microfilm and various forms of unitized microfilm such as microfiche, microfilm jackets, aperture cards, and chips are often employed. In any event, the microform portion of the system's work station includes a microfilm reader or copier that is mechanized to some degree. Communication with the computer portion of the system may be accomplished by either of two methods: One uses a remote terminal with a keyboard and possibly a CRT display; the other uses a punched card or punched paper tape equipment for sending messages to and from the computer. There is also equipment available that permits use of a single keyboard to communicate with both the microfilm and computer portions of the system. It employs a split viewing screen for simultaneously displaying information produced by both parts of the system.

If a person serves as the communication link between the computer and the microform storage unit, he is responsible for retrieving the appropriate micro-images upon receipt of the message from the computer. In other systems the computer message is used to automatically activate a microform reader that finds and displays the related micro-images for the user. In still another system, the computer message is used to control a mechanism that locates the proper microfilm image and makes a film-to-film copy of it.

The advantages of combined microform computer systems include an increase in the usefulness of the computer, reduction of computer storage costs, faster retrieval of information, and improved access to information. By using microforms to store close at hand large masses of previously acquired information along with current static or semistatic data and then using the computer to quickly identify the location of needed information and perform related ADP operations, new solutions are provided for both today's and tomorrow's problems.

The disadvantages are mainly that such systems usually require highly skilled designers and a rather substantial initial investment.

Special Considerations

It should be quite clear by now that microform systems do not offer a panacea for all of an agency's document dissemination problems. A cost-benefit study should always be made and pilot tests conducted before deciding to go ahead with a system. A major obstacle in any microform system is gaining user acceptance, and nothing should be left to chance. Appendix "E," (Department of the Air Force Regulation 12–40, March 5, 1971) provides a good example of the types of management controls required to insure the successful application of document miniaturization techniques.

When designing a microform system, serious consideration should also be given to capturing and maintaining key identifying data in machine language. Using source data automation techniques, this can be done for a small additional cost at the same time the labels are typed. The machine-language record should prove highly useful as a means for automatic preparation of finding aids, inventory lists, and new labels, and purging of the file.

Attention should also be given to subpart 101–11.5 of the Federal Property Management Regulations (41 CFR 101–11.5). While this regulation primarily applies to situations involving microfilming of permanent records in order that they can be destroyed, many of the safeguards provided therein should be observed in all microform systems.

The National Archives and Records Service, General Services Administration, operates microfilming service centers throughout the country. Government officials interested in these services or desiring assistance in microfilming and other paperwork management matters should contact the manager of the nearest GSA Regional Office or Federal Records Center.

IV. MANUAL NONCONVENTIONAL INDEXING SYSTEMS

The methods and equipment described in chapter III, "Microform Systems," were developed primarily to solve the problems associated with the physical handling and storage of documents. It was also explained in that chapter how microform systems can sometimes prove helpful in solving problems involved in looking up data in such voluminous listings as payrolls, directories, schedules, and price lists. If, for example, the user's problem is simply to look up the social security number, address, or telephone number of individuals with whom he deals, a microform system, or perhaps a conventional tool such as a printed directory or card file, is usually all that is needed.

If, on the other hand, retrieval involves searching for documents or information on the basis of subject topics or a variety of characteristics, attributes, or other features, the problem is quite a different one. The problems and limitations in using conventional methods and equipment in situations of this type are described in chapter I, "Why Information Retrieval Systems Are Needed." Chapter I, as well as chapter II, "How Coordinate Indexing Systems Work," explains how the nonconventional information retrieval systems may be employed to solve these problems. This chapter (IV) and the next one (V) describe the specific methods and equipment used in these nonconventional systems. This chapter covers manual methods and equipment, while the one that follows describes those employing mechanized equipment.

Manual nonconventional indexing systems, for the purpose of this handbook, include those where the search is conducted by manual methods. The tool or device may have been prepared manually, but some are, and most could be produced and updated by computers and other machines. Further, some of the tools could be converted to a microform format for ease in duplication and dissemination.

Types of Situations Where Nonconventional Indexing Systems Are Used

There are two basic types of situations where the methods and equipment in this and the next chapter are applicable. The first type involves organization of information mainly on the basis of *subject topics* for retrieval of textual documents or information. The second type is concerned with organizing information (data) on the basis of *characteristics or attributes* (also referred to as indexing terms in this handbook) for use in identifying and retrieving information or documents relating to individual people, places, or things. An example of this second type is a personnel skills inventory describing employees in terms of their education, experience, languages spoken, etc., for use in selecting people for promotion, reassignment, special projects, or other purposes. This second type of system is far less complex to design and operate than the first, mainly because it is relatively simple to develop and define the characteristics, attributes, or features to be used as indexing terms, while the task of selecting and defining subject topics is difficult and imperfect due to the ambiguity of the human language.

Prerequisites for a Successful Manual Nonconventional Indexing System

The most important prerequisite for a successful indexing system is *to obtain the right people for the job.* In all but the smallest and simplest of systems, special talents of two types are required. The first requirement is for the services of a skilled person to design the system and then return periodically to revise it, since there is no such thing as a finished design for an indexing system. If the system involves indexing documents by subject, the individual should have a thorough knowledge of both the subject matter field and indexing. If no such person is available, it may be necessary to use the team approach; that is, to bring together an individual who has a thorough

knowledge of indexing but only a limited knowledge of the subject matter with a person of the opposite qualifications.

The second type of talent needed is *qualified personnel to operate the system*. Again, if the system is used for indexing documents by subject, the indexers and searchers or indexer-searchers (and abstracters, if any) should have a thorough knowledge of the subject matter field and be properly trained in performing their duties.

Of next or perhaps equal importance is the need for *an operating manual or rule book*. The operating manual should include a vocabulary of indexing terms or a thesaurus, as it is commonly called, listing all indexing terms and defining how they are used in the system, supplemented by cross-references for synonyms and incorporating one or more devices for showing relationships among indexing terms. The operating manual should also include any other rules, guidelines, and reference aids needed for indexers, searchers, and users.

Another prerequisite for a successful system is *close coordination between the operators and users of the system* in all matters, including selection of documents or data entered into the system and continuous feedback on the effectiveness and value of the system. All users need to be kept informed about the new accessions, and new users should be oriented in regard to the contents and use of the system.

Another possible prerequisite, or at least desired feature of the system, is *compatibility with other systems* with which it may be interfaced now or in the future. This compatibility is of two kinds—system vocabulary and physical aspects. Today, it is seldom that any given collection of documents or data is of interest or value to a single organizational element. Somewhere within an agency, another agency, or the private sector, there is likely to be one or more groups of people collecting, storing, and retrieving similar if not identical information. System compatibility can therefore be of mutual benefit, possibly contributing through sharing arrangements to lower costs for all systems involved, while increasing the level of service to users.

Another important prerequisite is that there should be *a minimum of delay in entering the new items into the system* and making them available to the users. Not only should a search of the index reveal the presence of the item, but it should also be possible for the user to quickly obtain a copy of it.

Other prerequisites for providing good service to the user include ready access to the system and satisfactory performance of the system. A highly desirable but not necessarily essential feature would be that the system be readily convertible to an automated system.

Factors Affecting the Choice of the Type of Manual Nonconventional Indexing System

The major factors to be considered in choosing the most suitable type of manual nonconventional indexing system are as follows: (1) the present file size, growth rate, and estimated future size of the collection; (2) if the system is to be used for retrieving information by subject, the average number of indexing terms that will be assigned each document and the total number of indexing terms for the system; (3) if the system is to be used for retrieving information or identifying people, places, or things on the basis of characteristics, attributes, or features, the number that will be used to describe each item entered into the system; (4) physical form, format, cost, and source of the input; and (5) the extent to which the documents or data will have to be changed, updated, or deleted.

Other important factors to be considered in selecting the type of system include: (1) the average number of indexing terms to be used per search, the average number of searches per day, and the extent of workload fluctuations and peakloads; (2) the number and types of users and their physical location; (3) the physical form, format, and nature of the output required by the users; (4) service speed requirements; (5) special features required, if any, such as abstracting and evaluating documents and selective dissemination of information (SDI); (6) accuracy and reliability requirements; and (7) agency resources including availability of funds, personnel, and equipment for operation of the system.

Further information regarding the significance

CLUE-WORD EXTRACT CARD SYSTEM

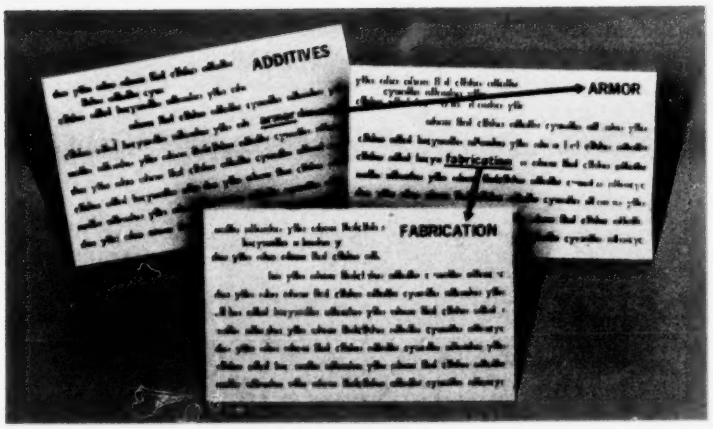


Figure 23

of these factors and guidelines on gathering the data, analyzing user needs, and selecting the right method and equipment are included in chapters VI and VII. A description of a number of systems employing manual nonconventional indexing methods and equipment is included in the records management handbook, *Information Retrieval Systems.*

Types of Manual Nonconventional Indexing Systems

The following are descriptions of the various types of manual nonconventional indexing systems, with a brief summary of the main advantages and disadvantages or limitations of each.

Clue-word Extract Card Systems. These systems are subject indexes consisting of 5- by 8-inch cards arranged alphabetically by "clue-words" (keywords) taken from the titles and text of the documents. Each card contains an extract of the document in which the keyword appeared. The extract is marked to indicate other keywords contained in the document, thus providing built-in "clues" as to other places to look in the file when conducting a search. Figure 23 illustrates how the "clue-word" principle operates. Information specialists, or preferably users of the system, evaluate incoming documents for relevancy. They underline the keywords in each selected document and place brackets around the portions to be extracted. They also assign additional indexing terms, if needed. Typically, tables of contents, author-prepared abstracts, and key illustrations are included in the extract.

Typists then prepare a 5- by 8-inch duplicating master containing the document number, title, author, other standard descriptive headings, and the extract with all keywords underscored. A sufficient number of cards are made of each document to permit the filing of one card under each of its keywords and the standard headings. Various colored cards, colored stripes, and corner cuts are employed to code the cards as to date, source, type of document, etc. The incoming material is maintained in a separate file.

The user begins his search by choosing a keyword he thinks should be helpful in identifying documents that may contain the information he is seeking. If, after scanning the cards filed under that particular term, he is still unable to find what he wants or needs further information, he takes note of other underlined keywords appearing in the body of the cards for "clues" as to where else to search for the needed information. He then refers to the other cards and thus proceeds with the search until he finds the desired information or until he has satisfied himself that the document collection contains nothing significant on the subject.

The major advantages of the clue-word extract card system are that no complicated input and output equipment is required; no preconstructed index vocabulary is needed (system is self-organizing); no special training is needed for conducting searches; it is highly browsable; and the extract cards are self-sufficient (it is usually not necessary to refer to original document). Further, this technique offers a simple, effective means for compacting text. The system concept is

SAMPLE PAGE FROM A PERMUTED (KWIC) INDEX

INCREASE IN QUALIFIED INVESTMENTS-LESS DEVELOPED COUNTRIES 0954 04 00
INVESTORS EXPENSES 0212 03 00
ED IN TRADE OR BUS., AND INVOL. CONVERSIONS - PROPERTY US 1251 00 00
INVOLUNTARY CONVERSION 1231 10 00
INVOLUNTARY CONVERSION -RECOGNITION- 1035 00 00
SS INVOLUNTARY CONVERSIONS UNDER SEC 10 0381 12 00
NTORIES INVOLUNTARY LIQUIDATION OF LIFO INVE 1321 00 00
SEC 482 ISSUES INVOLVED 0260 06 00
CLAIMS AGAINST U.S. INVOLVING ACQUISITION OF PROPERTY 1347 00 00
OVERASSESSMENT INVOLVING EXCESS PROFITS* 6214 01 04
OTHER THAN CONT IN THE IRC - ISSUES RELATED TO STATUTES 9999 42 00
SUSPENSION UNDER IRC 6501 -81* 9104 16 03
SAL OF COAL OR DOMESTIC IRON ORE DISPO 0272 00 00
TIMBER COAL OR DOMESTIC IRON ORE GAIN OR LOSS IN CASE OF 0631 00 00
IMPERFECT OR IRREGULAR ORGANIZATION 6012 03 02
ELECTION IRREVOCABLE 1361 02 00
RUSTEE OR BENEFICIARIES IRREVOCABLE TRUST - S IN HANDS OF T 1015 03 01
MUTUAL UNION OR IRRIGATION COMPANIES 0501 12 01
TAX INCURRED TO THE VIR ISL - S ON REDIST IN INC TAX L 0934 00 00
C ISL-GUAM-CAN ZONE-VIR ISL - ADM PUERTO RICO-TRST TER PA 4735 00 00
UERTO RICO-TRST TER PAC ISL-GUAM-CAN ZONE-VIR ISL - ADM P 4735 00 00
VIRGIN ISLAND RESIDENTS 0952 01 00
ISOLATION OF PROPERTY-GENERAL 2035 05 01
LATION ISOLATION-FEAR OF LOSS THROUGH SPECU 2035 05 06
ISOLATION-HAZARDS OF BUSINESS 2035 05 04
ISOLATION-MARITAL STATUS CONTEMPLATE 2035 25 10
ISSUANCE OF STOCK TO VOTING TRUSTEES 4321 04 01
LIMITATIONS OF ISSUANCE TAX 4502 01 00
BAILEY ISSUE -1970 COLF - 2042 00 00
FOREIGN CENTRAL BANK OF ISSUE FROM U.S. OBLIG & INC DERIVED BY 0945 00 00
ISSUE NOT PROPERLY PLEADED* 7453 13 02
STOCKH ISSUE RAISED VL NOT RAISED BY COMMIS 7453 15 04
JOINDER OF ISSUE* 7453 11 00
OBLIGATIONS ISSUED AT A DISCOUNT 0414 00 00
WHEN ISSUED TRANSACTION 1223 12 05
TAX ON ORIGINAL STOCK ISSUES 4501 01 00
SEC 482 ISSUES INVOLVED 0269 20 00

Figure 24

susceptible to application of computer techniques for information dissemination, automatic searching, and preparation of special finding aids. Therefore, anyone establishing a manual clue-word extract card system today should capture and retain the input data in machine language format for possible conversion to a computerized system at a later date.

The major disadvantages of the original clue-word extract card systems are the bulkiness of the files and the lack of a practical means for converting the file to an automated system.

Permuted Indexes. Permuted indexes are specially printed and organized printed manual indexes usually prepared by a computer from document titles, full text, a catalog, or index entries, as illustrated by the format of the KWIC (keyword-in-context) index shown in figure 24. There are various other formats, many of which are an improvement over this one. Some of the better known other permuted indexes are KWOC (Keyword Out of Context), WADEX (Word and Author Index), and SPINDEX (Special Permuted Index). To obtain these indexes a computer is programmed to alphabetically arrange the entries so that each document or other thing being described in the index is listed under each of its keywords.

KWIC indexes have been successfully applied in indexing operating procedures and directives, forms catalogs, the Controller General's decisions, and in numerous other situations. In cases where permuted indexes are used for indexing procedures and directives, a special dividend may be expected—the index will highlight inconsistencies, duplications, and omissions. With the increased usage of permuted title indexing, authors are giving more attention to selecting meaningful, useful titles; and this, together with the improved formats and low costs, is enhancing the use of permuted title indexing. The retrieval capability of permuted indexes can be increased by inclusion of additional indexing terms selected from an index vocabulary such as the *Thesaurus of Engineering and Scientific Terms* used by the Department of Defense and other Government agencies.

The major advantages of permuted indexes are the following: (1) the relatively low overall cost (in some situations the index can serve as a low cost substitute for manually prepared indexes or can make it practical to provide an index where none existed before); (2) speed and ease of preparation (computer printouts that serve as final copy for offset printing of the index can be obtained in a matter of hours); (3) ease of revision (the speed of a computer makes it possible to print out an entire new index including any revisions, rather than trying to manually patch up a printed copy as revisions are made); (4) more meaningful and browsable than conventionally printed indexes (the one or two word entries do not normally provide an entire concept); and (5) reduction in the time required to announce new documents and enter them into the system.

Permuted indexes can also serve as a means for developing in-house capability in the use of computers for information processing, and in at least some instances will result in the establishment of a computer data base that may serve even more important purposes in the future.

The major disadvantage of the permuted index is that it does not provide cross-references for synonyms; therefore, it is subject to searching problems created by the author's inconsistencies in word usage and the normal ambiguity of human language. Further, if limited to document titles only it becomes a shallow index; if applied

to the entire text, it may become too cumbersome to be practical.

An ideal permuted index for procedural manuals and similar publications would include a permuted listing of titles for the parts, chapters, sections, the paragraph or other headings, and any abstracts or other summaries of the contents of the documents.

Columnar Card Systems. These systems, as shown in figure 25, are coordinate indexes in which one card is prepared for each indexing term used in the system. The numbers of all documents indexed under each term are entered on its term card. Each term card is divided into ten columns, 0 through 9, and the document number is posted in the column corresponding to its terminal digit. Searches are conducted by selecting those term cards that seem pertinent, and then matching document numbers column by column to locate any numbers that appear on all the cards. The cards are usually prepared and maintained manually, either by hand or typewriter; the basic data, however, could be maintained in machine language form and the cards produced by a computer.

The major advantages of the columnar card systems are that the costs for supplies and equipment are extremely low; they permit parallel searching of the index file (rather than requiring a card-by-card serial search); and they are simple and easy to maintain and use, being highly manipulative and browsable.

The major disadvantages or limitations of the columnar card systems are that it is usually necessary to refer to a second document, such as an abstract or even the document itself, to obtain a document description or to determine a document's relevancy; and if the system is used extensively, searching can become slow and tedious should the columns of numbers become long and individual searches involve several indexing terms.

Dual Dictionary Systems. These systems, illustrated in figure 26, are similar in design and use to columnar card systems, except that all the indexing terms and document numbers are printed on two identical lists mounted side by side in a binder. Instead of matching cards during the search process, the user looks up the first term

SEARCHING WITH COLUMNAR CARDS

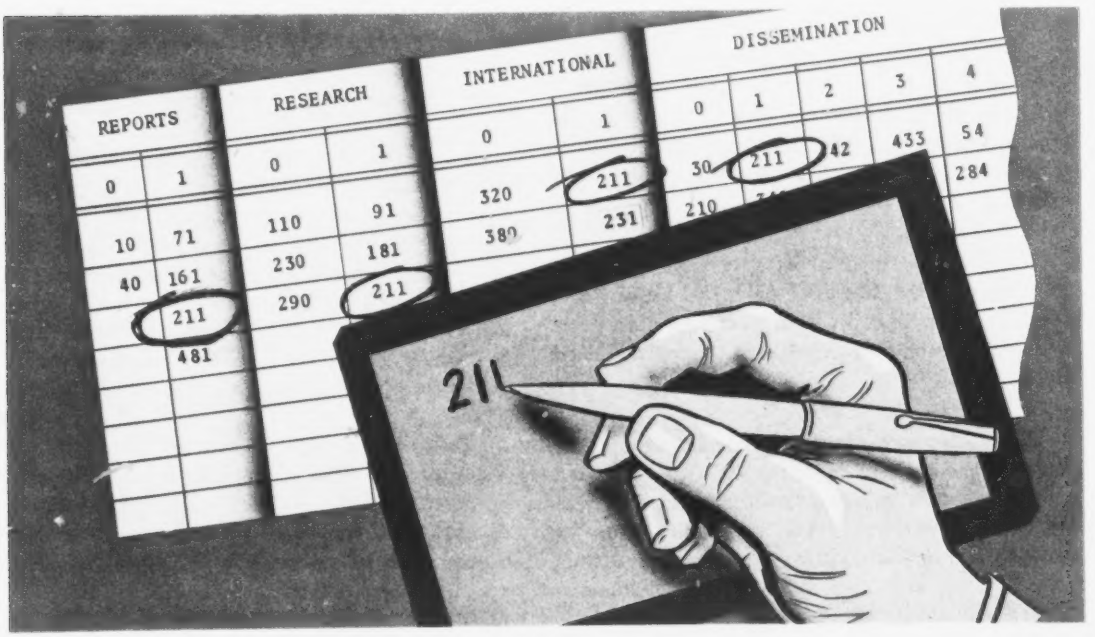


Figure 25

SEARCHING WITH A DUAL DICTIONARY

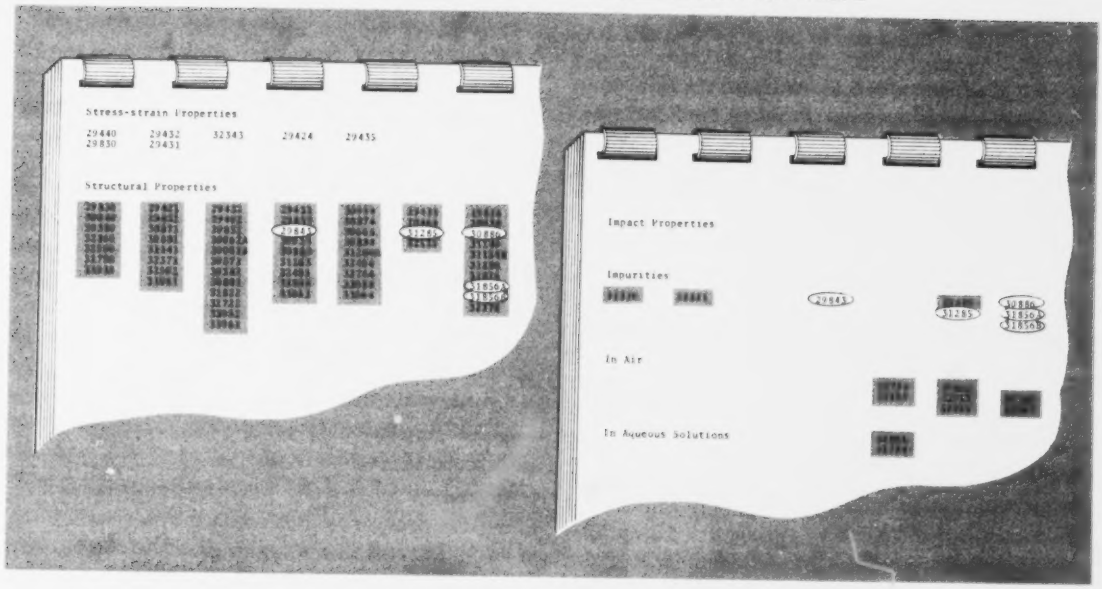


Figure 26

in its alphabetical location on the left side of the dual dictionary and then locates the second and other terms on the right side (or vice versa), checking for coinciding numbers at each step until the search is completed. Usually many copies of the dual dictionary are made and distributed to individual users.

The dual dictionary is best suited to those situations where there are many users in different locations. The dictionary's usefulness can be increased by furnishing with it abstracts of the documents and a copy of the thesaurus or other vocabulary of indexing terms.

The data for the dual dictionary may be manually maintained; however, more often it is maintained and updated by computer and then periodically printed out, duplicated, and distributed to the users.

The major advantages of dual dictionary systems are the same as those for the columnar card system, plus an important, additional one—these systems permit numerous individual users or groups of users to do their own searching, thus reducing the workload at the main information center and giving the user direct access to the system.

The disadvantages of the dual dictionary are also essentially the same as those for the columnar cards, with one exception—they are far more costly to maintain; however, if the number of users is sufficient the overall systems costs could, by comparison, be relatively low.

Edge-notched Card Systems. These systems, as illustrated in figure 27, are cards containing punching positions, represented by pilot holes along one or more of their edges, used in recording in coded form such data as indexing terms, dates, and numbers. The data is recorded by punching out the area in front of the pilot hole. The edge notching may be done manually by a hand punch or semiautomatically by special equipment. The interior of the cards, which are printed in various sizes and formats, may be used for written information or graphics. Typically, one card is prepared for every document or item being indexed.

To search the file, needles are passed through the appropriate pilot holes in the deck of edge-notched cards. The selected cards (those that are notched) fall out, while the others remain on the needle. Searching usually involves numerous needle passes. Other devices and equipment, in addition to the standard needles, are available for assisting in the search process.

EDGE-NOTCHED CARDS

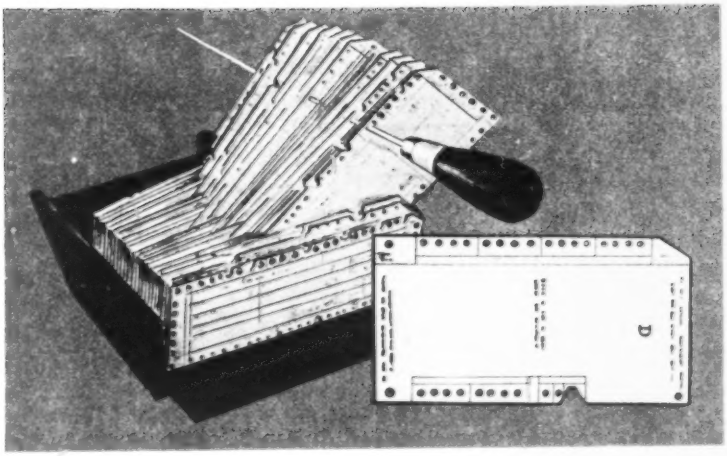


Figure 27

The major advantages of the edge-notched card systems include low cost, simplicity, the ease with which users may browse, immediate access to the description of the documents or things involved in the search process, and in many situations, elimination of the need to maintain the cards in a precise sequence.

The major disadvantages of the edge-notched card systems include limitations on the amount of coded data that may be recorded on the card; slowness and awkwardness in the search procedure if the cards are used extensively for complex searches (due to the system requirement of serial searching); limitations on the size of file (many information specialists consider 5,000 cards to be the upper practical limit); the somewhat complicated code patterns; and the possible difficulty in detecting coding (edge-notching) errors.

OPTICAL COINCIDENCE CARDS AND VIEWER

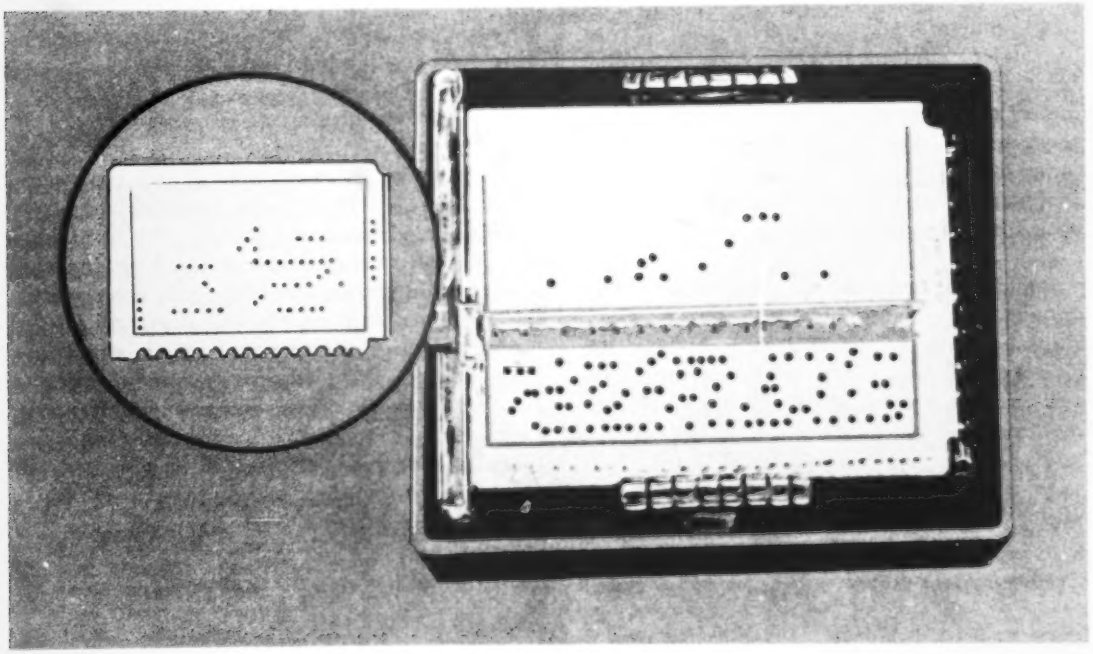


Figure 28

Optical Coincidence Systems. These systems, as illustrated in figure 28, employ cards (or sheets) with a fixed number of dedicated positions or address locations for drilling (or punching) holes representing the individual documents or items being indexed. A separate optical coincidence term card is maintained for each indexing term. After each incoming item has been indexed and assigned a serial number or optical coincidence card address location, all related term cards are removed from the file and machine drilled or punched in the appropriate position.

Searching is accomplished by first selecting the optical coincidence term cards pertinent to the query. The selected cards are then stacked and are placed in front of a light source to visualize the existence of coinciding holes. The position of the matching holes on the cards indicates the number or address locations of any documents or items that fully satisfy the search question.

In addition to identifying documents or other items pertinent to a query, the cards may also be used as a data manipulation and tallying device for compiling statistics; or, through the use of transparent overlays, as a means for presenting statistical data in a visual manner. Although in most optical coincidence systems the cards are drilled, manipulated, and interrogated manually, there is equipment available for machine controlled drilling of the cards, machine counting of holes, and automatic printout of numbers. In the system developed by the National Bureau of Standards, the user can see an enlarged microfilm image of the related document abstract during the interrogation process.

The optical coincidence cards most commonly used are about 9 inches in size and can accommodate up to 10,000 documents or items and 1,000 indexing terms. Prescored punched cards that can accommodate 480 items are also sometimes used.

The major advantages of optical coincidence systems are manipulatory ability; encouragement of browsing by the user; rapid searching speeds (partly because these systems permit parallel searching of the index file rather than requiring a serial card-by-card search); low cost for supplies and equipment; simplicity; and fast, easy read-out of the search results.

The major disadvantage of optical coincidence cards is that it is usually necessary to refer to a second information source to obtain a description of the document or item, or to determine its relevancy. Another possible problem is in error correction; however, some types of input equipment help keep errors to a minimum by preventing redrilling in the same hole.

Special Considerations

This chapter reveals that there are many simple, rather inexpensive nonconventional indexing systems which, although manually operated, offer significant advantages over conventional systems for organizing and retrieving information. In many situations today, one of these manual systems may be all that is needed to solve the information retrieval problem. However, in most situations it will some day become desirable to convert the system to one that takes advantage of computer capabilities for maintaining, reorganizing, reformatting, merging, updating, and purging of information in the file, and manipulating, selecting, and presenting the information.

In order to do these things the data contained in the index file must be in machine language. Consequently, when developing and installing any manual nonconventional indexing system, serious consideration should be given to recording the index data in machine language as a by-product of the input operations. Such devices as paper tape and magnetic tape or card typewriters are ideally suited to this purpose. Further, as mentioned earlier in this chapter, the machine language data base, with the aid of a computer, can be used to produce many of the nonconventional manual indexing tools.

V. NONCONVENTIONAL MACHINE INDEXING AND RETRIEVAL SYSTEMS

The significance of nonconventional machine indexing and retrieval systems rests not in the number of basic types of equipment that are available, but in the wide variety of tasks these systems can perform, their flexibility, and their future potential. In numerous instances the indexing, storage, and retrieval operations are, or could be, a satellite of a larger integrated automatic data processing (ADP) system. Today, there are many instances where the data base maintained for an ADP system could, with slight modification and expansion, serve as the nucleus for a highly useful information retrieval system. On the other hand, there are situations where machine nonconventional indexing and retrieval systems could largely pay for themselves by solving logistical and other problems involved in the preparation, stocking, distribution, replenishing, and control of documents.

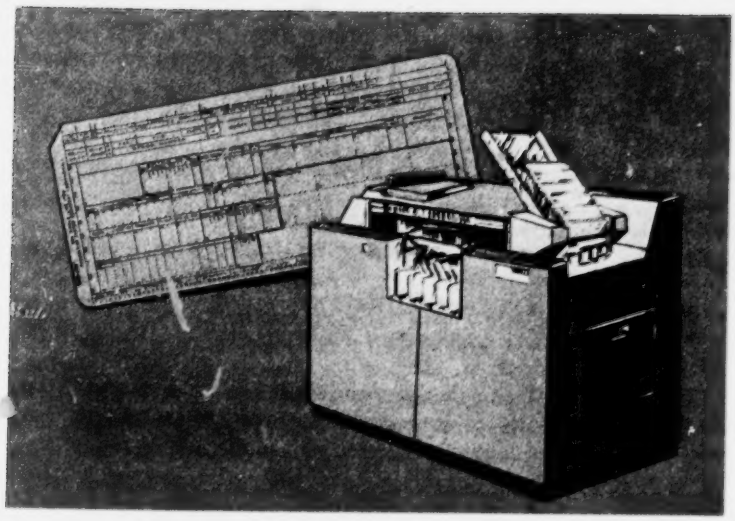
Obviously, the equipment used in machine nonconventional indexing and retrieval systems is usually more expensive than that used in the manual systems. Further, the machine systems are generally more difficult to design and operate. However, these conclusions can be misleading, and in practice they prove to be but a slight barrier in installing a machine system. The first reason for this is that instead of acquiring your own equipment, you could more than likely obtain machine time on equipment already installed in the agency or available through a service bureau. The second reason is that there are available many standard and special machine programs (machine instructions and procedures) that, with slight modifications, can be adapted to the job at hand. When one considers these possibilities, and the indisputable move toward automation in all areas, it becomes increasingly clear that any information retrieval system study should include a thorough investigation of machine methods for doing all or part of the job either now or in the future.

Types of Situations Where Machine Indexing and Retrieval Systems Apply

There are two basic situations where the methods and equipment described in chapter IV and in this chapter may apply. In the first situation, i.e., *retrieval of textual documents or information on the basis of subject topics,* machine systems are proving highly satisfactory; and in addition, many of the systems can automatically furnish the user with a complete description of the document or permit him to view the document or—perhaps immediately—to obtain a copy of it. In the second type of situation, i.e., *retrieval of information or documents on the basis of characteristics or attributes,* machine systems have the additional capability of being able to automatically retrieve selected data about a person, place, or thing, or a complete description or image of it. There is also an additional type of situation where only nonconventional machine information retrieval systems apply—*the storage and retrieval of large masses of data in what are commonly called data banks.* Machine methods and equipment can be used to update these files, to automatically and selectively transfer data from one file to another, and, on demand, to selectively retrieve data and perform data manipulations.

Prerequisites for a Successful Machine Indexing or Retrieval System

All the prerequisites cited in chapter IV for a successful manual nonconventional indexing system are also important to the success of machine systems, and therefore should be carefully noted. An additional prerequisite for machine systems is the ready availability of personnel, either on a full or part-time basis, who are trained and experienced in the operation of the equipment. Another important prerequisite is the accessibility of equipment being able to have access to it at the right time and frequency required by the users.



EAM PUNCHED CARDS AND COLLATOR

Figure 29

Another important but not necessarily essential feature is that the data elements and codes be compatible with other computer data banks in the same field of interest so if it should later become necessary or desirable the data can be readily exchanged, compared, or combined on a machine-to-machine basis.

Factors Affecting the Choice of the Type of Machine Indexing or Retrieval System

In addition to the factors cited for manual systems in chapter IV, which also apply here, machine systems are concerned with machine record lengths. Machine record lengths involve: the number of data elements (for example, date of birth) per record; the number of data items (for example, year of birth) within the data element; and the total number of characters (alphabetical, numerical, and special) per record.

Types of Machine Indexing and Retrieval Systems

The following are descriptions of the various types of machine nonconventional indexing and retrieval systems, together with a brief summary of the main advantages and disadvantages or limitations of each.

EAM (electrical accounting machine) Punched Card Systems. These systems employ cards divided into vertical columns, with each column then divided into 12 punching positions. Each column can be used to record, by means of one or more punched holes, a single alphabetical, numerical, or special character. The cards are divided into segments (fields) of various lengths for recording such individual data elements as the following: titles, segments of text, names, dates, addresses; and code numbers representing names of organizations, forms, products, or indexing terms. A wide variety of equipment is available for punching, sorting (including electronic high-speed sorters), collating, interpreting (card printing), selecting, and analyzing the punched cards, in addition to equipment for performing arithmetic operations and preparing printed listings. Figure 29 illustrates a punched card and a special collating machine.

Punched card systems were originally intended for use in performing statistical and accounting operations. In using punched cards as a medium for recording and retrieving data for information retrieval, the system designer has to adjust his methods to the capabilities and characteristics inherent in punched card equipment.

In organizing a punched card file for a coordinate index, there are two general ways for recording the index data and arranging the punched card file. One way is to prepare one or more punched cards, as needed, for each document or other thing being indexed and record thereon a

limited amount of data identifying the document, plus all the assigned indexing terms (subjects, characteristics or attributes). The file is arranged in document number sequence. The second way to organize the index file is to prepare a separate punched card for each indexing term assigned each document. Each card usually contains only the document number and the assigned index term; the cards are arranged in groups according to the indexing terms. This is commonly referred to as an inverted file.

The first way of organizing the file has the disadvantage of making it necessary to pass the entire punched card file through the equipment each time a search is conducted; however, it has the advantage of furnishing the user at least a brief description of the document. The second approach has the advantage of making it necessary to process only those punched cards representing the indexing terms involved in the search, which is conducted by comparing the punched cards representing any two of the indexing terms to determine coinciding document numbers, and repeating the matching process for the remainder of the term cards involved. This second method has the disadvantage of providing the user with the document numbers only, thus making it necessary for him to refer to a second source or to the document itself to obtain a description of the document and determine its relevance to the search question.

Another method of recording the indexing terms on the punched cards is to use superimposed coding, which offers greater data compaction but requires considerably more skill on the part of the system designers and operators.

The major advantages of punched card systems, when used for information retrieval, are their ease of manipulation; their relative simplicity (when compared with computers); their ease in reformatting, transferring, extracting, updating, and duplicating data; their capability for producing low-cost duplicate sets and printed listings; the ability of the cards to also be manually selected, read, and refiled; and their ready convertibility to computer systems.

The two major disadvantages of punched cards used as information retrieval systems are (1) the relatively slow searching speeds and the resulting slowness of the entire process of conducting coordinate-type searches; and (2) the limited accessibility of the punched card system, including a restriction upon the freedom of the user to browse, due to the fact that card files and equipment are usually maintained in a machine room and their use requires trained machine operators.

Most systems employing punched cards for coordinate indexing consist of less than 20,000 cards; however, if used primarily for simple data lookups and only occasionally for coordinate-type searches, a file of 50,000 or more may be feasible. Therefore, punched cards, due to this reason and the advantages described above, are particularly well suited to personnel skills inventory and other systems that usually entail a large volume of manual data lookups and recurring or special printed listings of various types and formats, but only a limited number of coordinate-type searches. Punched cards may also be used for selective dissemination of information (SDI) systems, but since today computers are more often used for this purpose, selective dissemination of information systems are included in the latter category.

Computers. Computer equipment is of two basic types: analog and digital. Analog computers may be likened to a slide rule or an automobile odometer, since they work with physical quantities and compute by measuring. Digital computers, on the other hand, work with numbers or digits and compute by counting. Digital computers are divided into two classes, special and general; computers in the general class are normally used for automatic data processing (ADP) and information retrieval. A typical equipment configuration is shown in figure 30.

Computers are the most versatile and powerful of all the devices used for information retrieval, due to their high processing speeds, accuracy, ease of updating, ability to perform complex transactions automatically and to communicate with each other, and their ability to provide the user with a wide range of on-line search capability and off-line services and tools, including permuted indexes such as the KWIC index described in chapter III. Another advantage offered by the computer used for information retrieval purposes is its usefulness for administrative and

COMPUTER EQUIPMENT

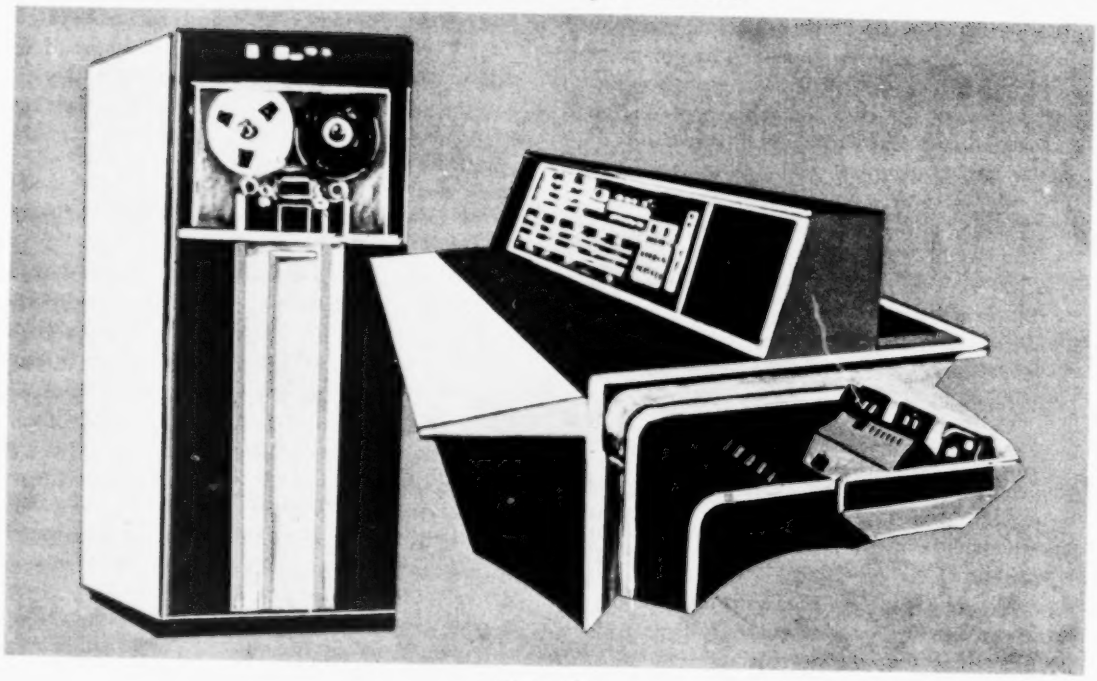


Figure 30

logistical tasks. For example, it can be used to prepare requisitions and announcements of new accessions, to operate a selective dissemination of information system (SDI), to bill for user charges, and to maintain an inventory. These special tasks are all accomplished as a by-product of normal input and output operations. The computer can also be useful in controlling access to restricted or classified information.

Three of the major limitations in using the computer for information retrieval are (1) high input costs; (2) shortage of systems analysts and programmers having experience in information retrieval systems; and (3) lack of low cost, on-line computer mass memories. Solution to the input problem depends on applying source data automation (SDA) techniques, including capturing data in machine language as a by-product of other processing operations and using optical character recognition (OCR) equipment for automatic document reading and conversion to machine language.

The problem of the scarcity of experienced systems analysts and programmers in the information retrieval area is still acute, and the only significant relief available at present is to utilize existing computer programs and operating systems developed and designed by others. The problem of developing low cost, on-line mass memories is the object of intensive research by many computer manufacturers and others, and while the results look promising, none are yet commonly available.

Notable progress has been made in computer-user communications. While most systems still require the preparation of a punched card to gain access to the computer and most of the output is still in the form of printed forms and listings, punched cards, or microfilm produced by COM equipment, there are more and more systems that permit direct communication between the user and the computer.

These two-way communications are accomplished by means of remote terminals employing teletypewriters, other types of typewriters, and cathode ray tube (CRT) devices with keyboards and light pens, as illustrated in figure 31. By keying in the proper user identification code and

following a prescribed routine for interrogating the computer, the user is able to obtain answers to his questions or possibly update, edit, or delete data in the computer store. With the addition of the light pen, he is able to pinpoint numbers, words, or phrases appearing on a CRT to make searching easier and faster or to quickly instruct the computer to delete, change, edit, or transfer stored data. Data in the computer store can also be used to produce charts and other graphics.

Significant refinements in computer programs, which make communication with the computer more like conversation, plus improvements in the hardware and reduction in equipment costs, assure that the remote terminal will eventually become commonplace. Since the main use of the remote terminal is to retrieve and manipulate data, those who manage the agency's records and other information resources can expect an increasing demand on the part of management to computerize the agency's important data bases, particularly those that are dynamic in nature.

Rather than describing computers in accordance with their size, type, or operating characteristics, this chapter describes them in terms of the ways they are most often used for information storage and retrieval.

Computer index searching systems are those used to search index files where the indexing itself is performed manually. Indexers, using a guide such as a thesaurus of indexing terms, assign the indexing terms to the individual documents. The indexing terms are then usually coded, that is, converted to a numerical representation, and along with other pertinent data recorded in machine

A CRT TERMINAL WITH A LIGHT PEN

Figure 31

language by means of a keyboard device such as a card punch, paper typewriter, or other encoding device. The output of the machine language recording device is made a part of the computer index file.

As in the case of punched card systems, the index file can be organized and arranged in either of two ways—by document numbers or by indexing terms. If the index file is arranged by document numbers, the file description of the document may include the title, author, date, and a list of indexing terms assigned to the document, together with other bibliographic data and possibly an abstract or extract of the document. If the index file is organized and arranged by indexing terms, only the number of each indexing term and the numbers of all documents assigned that term are shown on the main computer's index file (inverted file arrangement). In systems arranged by indexing terms, a separate auxiliary bibliographic record similar to the main index record for systems arranged by document numbers, is often maintained on the computer.

When conducting a coordinate-type search in those systems where the index file is organized and arranged by document numbers, it is necessary to make a serial search of the file, which may necessitate the loading and unloading of several reels of magnetic tape if the information is stored on tape. Whenever a document satisfies the search requirements, its complete description is immediately available.

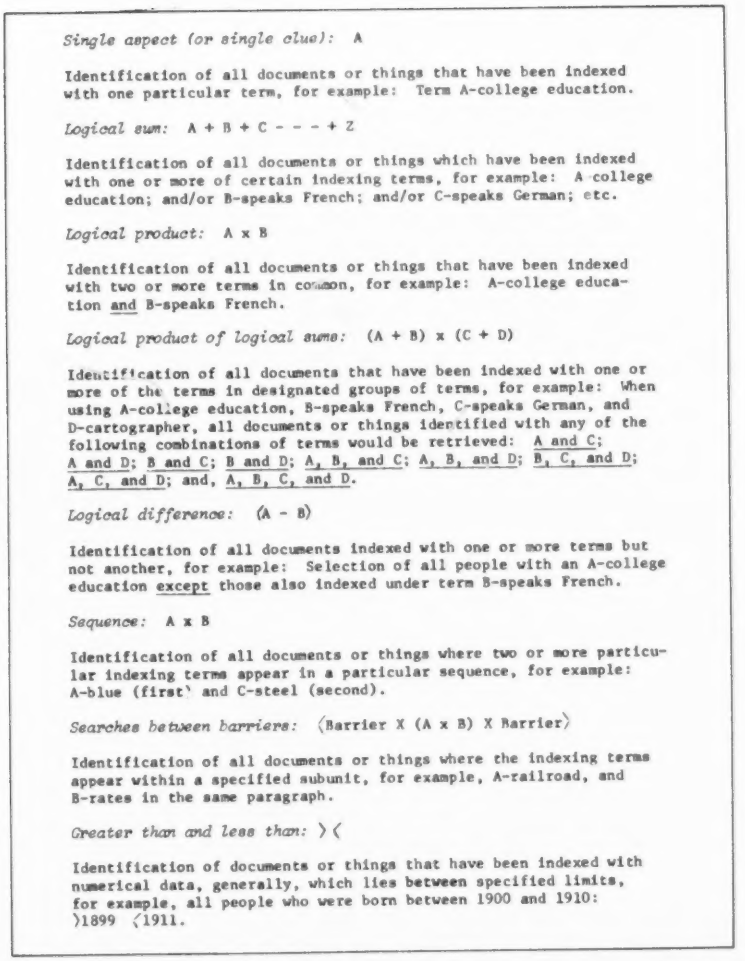
When conducting a search in those systems where the main index file is organized and arranged by indexing terms (inverted file) the entire index file, which is highly compact, can often be quickly searched on-line. However, it is necessary then to go to the auxiliary computer index file or perhaps a separate manual index file or the document itself, to obtain the description of the document. Determination as to which file arrangement is best is governed by such factors as the index file size, the number and frequency of searches, the type of equipment and machine program used, the needs of the users, and the capability of the computer to conduct more than one search at a time.

In addition to the general advantages of the computer mentioned earlier, its use for index searching can be beneficial in numerous other ways. The computer can be used to provide statistics on the frequency of use of indexing terms in both indexing and searching and the frequency of association between indexing terms—information that will provide valuable clues in system modification and control. The computer can be used to construct or prepare the index dictionary or thesaurus of indexing terms and the various special reference aids for indexers, searchers, and users.

Computer automatic indexing and searching or "full text" systems substitute the computer and its programmed instructions for human effort, not only in conducting searches but also in indexing documents. The full title, the full text, and other bibliographic data including an abstract, if any, are converted to machine language for input to the computer. Automatic indexing is based on the general principle that the noncommon words in the document are suitable indexing terms. In order to make it possible for the computer to choose the noncommon words, it is supplied with a list ("stoplist") of such common words as "the" and "of," which are not to be included in the index. In the input processing the computer compares each word in the text against those contained in the stoplist, and where they do not match the word becomes an indexing term. Typically, in deriving an index in this manner each document, paragraph, sentence, line, and word is automatically assigned a serial number and the computer index file is arranged in concordance fashion. Following each of the indexing terms (the noncommon words), the serial number is listed for each location where the term appears in the text. In addition to the index, the complete original text is also usually maintained in machine language.

Numerous techniques are used for conducting computer searches of the full text index file. Typically they include the Boolean algebra or set theory concepts employing the computer logic operations of and (intersection), or (union), and but not (negation), as illustrated in figure 32. Additional techniques commonly employed include specifying how many times the indexing term must appear in a document (word frequency counts) and the proximity of one indexing term to another. Further refinements in searching may be

CONDUCTING SUBJECT-TYPE SEARCHES BY COMPUTER



Single aspect (or single clue): A

Identification of all documents or things that have been indexed with one particular term, for example: Term A-college education.

Logical sum: A + B + C - - - + Z

Identification of all documents or things which have been indexed with one or more of certain indexing terms, for example: A-college education; and/or B-speaks French; and/or C-speaks German; etc.

Logical product: A x B

Identification of all documents or things that have been indexed with two or more terms in common, for example: A-college education and B-speaks French.

Logical product of logical sums: (A + B) x (C + D)

Identification of all documents that have been indexed with one or more of the terms in designated groups of terms, for example: When using A-college education, B-speaks French, C-speaks German, and D-cartographer, all documents or things identified with any of the following combinations of terms would be retrieved: A and C; A and D; B and C; B and D; A, B, and C; A, B, and D; B, C, and D; A, C, and D; and, A, B, C, and D.

Logical difference: (A - B)

Identification of all documents indexed with one or more terms but not another, for example: Selection of all people with an A-college education except those also indexed under term B-speaks French.

Sequence: A x B

Identification of all documents or things where two or more particular indexing terms appear in a particular sequence, for example: A-blue (first) and C-steel (second).

Searches between barriers: (Barrier X (A x B) X Barrier)

Identification of all documents or things where the indexing terms appear within a specified subunit, for example, A-railroad, and B-rates in the same paragraph.

Greater than and less than:)(

Identification of documents or things that have been indexed with numerical data, generally, which lies between specified limits, for example, all people who were born between 1900 and 1910:)1899 (1911.

Figurre 32

achieved by placing special conditions on the search, such as that the index term must follow the phrase 'in conclusion," or must appear in the first sentence of a paragraph, and so on.

The United States Air Force Legal Information Thru Electronics (LITE) system at Denver, Colo., available for use by all Government agencies, is a good example of the versatility of an automatic indexing and searching system. The LITE system includes the full text of all published Decisions of the Comptroller General of the United States; Armed Services Procurement Regulations; Air Force manual 75-34, Reporting of Transportation Discrepancies in Shipments; and some 30 other sets of documents. When requesting a search the user has three choices as to the output: A list citing the documents found to be pertinent to the search question; a three-line KWIC listing from those parts of the document text where the index term appears; or a complete printout of the full text of the documents.

By using many of the same techniques as those employed for automatic indexing and searching, computers can also be programmed for development of classification systems, automatic classification of documents, and automatic preparation of abstracts and extracts. However, work in these areas is largely experimental. Other forms of automatic indexing include techniques employing statistical word counts and association maps. Work has also been done in refining automatic indexes by adding a thesaurus-like computer record that is used to provide guidance and assistance in either the indexing or searching process.

No system for indexing textual material by subject is without its faults. All things considered,

a well designed and properly operated computer indexing and searching system can be expected to perform about as well as those information retrieval systems where the indexing is done manually.

The major limitations of automatic indexing, searching, and preparing abstracts or extracts are the cost and the high degree of expertise required to design and operate such systems. However, the cost factor will become less critical as more and more offices move toward integrated information processing and retrieval systems that ultimately may include such features as computer-assisted document preparation and revision, computerized editing and preparation of the table of contents and index, and computerized printing. A copy of the same computer magnetic tape that goes to the Government Printing Office for use in automatic photocomposition and printing or is used to produce microform copy by COM equipment will also serve as input to the automatic indexing system, thereby eliminating one major cost—that of conversion of the information retrieval system input to machine language. These integrated information processing systems have one advantage that for many organizations may be far more important than the possible savings in cost—namely, the reduction in the period that elapses between the time an important event occurs, a fact is discovered, or a decision rendered, and the time the information is in the hands of those for whom it is destined or who may be searching for it.

Those persons interested in learning more on the subject should read NBS Monograph 91, Automatic Indexing: A State-of-the-Art Report, reissued February 1970 by the National Bureau of Standards (NBS) U.S. Department of Commerce.

Selective dissemination of information (SDI) systems are those that employ the computer or punched cards to provide individual users or user groups with tailormade announcements of new documents in their individual spheres of interest. The user's interest profile may be developed by having him look over the thesaurus of indexing terms and select those terms that reflect his areas of interest. The results are then recorded on a magnetic tape. Each time a new document is indexed, the indexing terms assigned the document or appearing in the abstract are compared with those stored on the user profile magnetic tape. In those instances where the requirements for a match are satisfied, the user is sent an announcement of the document, including its abstract, if any. Figure 33 illustrates an article announcement (abstract) card and a card used by the recipient to respond to the SDI system operators. Note that there are blocks on the recipient's response forms for him to use in indicating whether or not he wants to see the document and if not, why not, thereby providing the system operators with the necessary feedback.

An interesting variation of the SDI technique is to develop interest profiles for major projects or programs, instead of for people, and to use the computer to keep the project director informed of any new documents on the subject.

While the costs for SDI systems are appreciable, the costs may not be considered unreasonable from management's point of view, particularly in the areas of scientific and technical research and development. However, scientists and engineers are not the only professionals having problems in wading through the tremendous volume of new documents made available to them, while at the same time trying to make sure they have not missed any documents that could have a major impact on their work.

The trend toward using group interest profiles rather than the profiles of individual users is resulting in less expensive and many times more practical SDI systems. SDI systems are especially valuable in providing the user with "peripheral vision" of information of direct interest to him, but which might be overlooked without the benefit of an SDI service.

Computer data storage and retrieval systems, sometimes referred to as data banks, are those used to store, retrieve, and manipulate large volumes of data (facts, numbers, letters, and symbols representing basic elements of information that can be processed or produced). Data bases may be either of two types or perhaps a mixture of the two: (1) recurrent or dynamic data, which is subject to change, and (2) noncurrent or static (archival) data relating to a unique event or representing an unchanging situation. The data base

SAMPLE SDI SYSTEM NOTIFICATION CARDS

5473

RESNICK A
RELATIVE EFFECTIVENESS OF DOCUMENT TITLES AND ABSTRACTS FOR
DETERMINING RELEVANCE OF DOCUMENTS
IBM ASDD YORKTOWN HGTS NY, 17-033, OCT 1961

INDIVIDUALS WHO RECEIVED DOCUMENTS THROUGH A SELECTIVE DISSEMINATION OF INFORMATION SYSTEM WERE ASKED TO DETERMINE THE RELEVANCE OF DOCUMENTS TO THEIR WORK INTERESTS ON THE BASIS OF TITLES AND OF ABSTRACTS. THE RESULTS INDICATE THAT THERE WAS NO SIGNIFICANT DIFFERENCE BETWEEN THE USEFULNESS OF TITLES AND OF ABSTRACTS FOR THIS PURPOSE. 2 PAGES

Article (Abstract) Announcement Card

H P LUHN (NAME) 923 (DEPT) 800 (LOCATION) DATE MAR 8 5473 (DOC. NO.)

FROM SDI SYSTEM IBM ASDD YORKTOWN HEIGHTS, N.Y.

INSTRUCTIONS:
1. Read the Abstract
2. Punch the Appropriate Box
3. If you care to comment punch the comment box and write your comments on this card
4. Return this card to SDI

Of Interest, Document Requested
Of Interest Document Not Wanted
Of Interest, Have Copy
Of No Interest
Comments

Recipient's Response Card

Figure 33

may be specially created for information retrieval purposes, as in the case of weather data, or it may be used to serve multiple purposes. For example, census data is used for developing statistics and preparing reports as well as for information retrieval.

The social security and Federal income tax data bases are used mainly for automatic data processing purposes and only secondarily for information retrieval. Computerized management information systems also serve two purposes—to automatically produce reports and other communications and for information retrieval. It is the exception rather than the rule that a data bank is created and used solely for information retrieval. However, unless careful attention is given to the information retrieval needs in the planning and design of these multipurpose computer systems, there may be serious limitations or problems when later attempts are made to use the system for retrieving information.

For example some of the earlier ADP systems, in attempting to keep the machine record as short as possible, omitted such important data as the names of the individuals whose records were being maintained in the computer. Others were designed in such a way that individual items of data could not be selectively retrieved because the data was merely printed out in long lines without column headings. Sometimes the data was expressed in coded form, making it necessary for the user to refer to a special table to interpret the printout. Another problem, which is particularly critical at this time, is the lack of standardization or compatibility in data elements, thus making it difficult and sometimes impossible to exchange, compare, or combine data maintained in separate systems but relating to the same people, places, or things.

Unlike computer index searching systems and computer automatic indexing and searching systems, computer data storage and retrieval sys-

tems files are organized and arranged in an endless variety of ways. Generally, the method used initially for organizing and arranging the data prior to conversion to a computerized system is also the method selected for the new system. Thus, computerized census records are organized and arranged on a geographical basis much as they were before the advent of the computer. Personnel data banks are usually organized by the name or identification number of individual employees or job applicants. However, the computer offers one distinct advantage not normally possible or practical in conventional systems—the capability of organizing and arranging the same data in a variety of other ways. For example, personnel data can, in addition to the basic arrangement, be organized on the basis of organizational assignment, position classification series, years of service, etc., for direct searching or preparation of special listings.

Case files (files organized by the names or identifying numbers of people, places, or things) represent approximately 85 percent of the folderized records of the Federal Government. These files contain a wealth of data, but when stored in conventional systems the data is buried so deep in the file that it receives only limited use. By converting the data in these files to computerized systems, it becomes possible to readily select, extract, compare, and manipulate the data in an endless variety of ways to meet day-to-day operational requirements, to provide statistical data for management decisions, and to satisfy unpredictable needs of the future.

The only serious disadvantage of computer data storage and retrieval systems at present is their cost. However, the cost picture is gradually changing due to reduction in computer input costs through the application of SDA techniques; larger and cheaper computer data storage devices; faster processing speeds; and faster, less costly methods and equipment for retrieving and producing the system output.

Tomorrow's records manager will more than likely discover that most of the data needed to satisfy his clientele will be available via the computer and that his conventional files will serve mainly as depositories for selected original documents having legal or archival value. Today's records managers should therefore survey every existing record series for the purpose of identifying those which at some future date will or should be converted to a computerized data base and then work with management in developing an orderly schedule for the conversion.

Other Machine Indexing and Retrieval Systems

While most of the microform equipment described in chapter III is designed primarily for storage of documents or data in miniaturized form, some also have the capability to conduct logic-type searches. These are as follows:

Motorized (mechanized) Roll Microfilm with Photo-optical Binary Code. Although retrieval speeds with this type of equipment are not nearly so fast as those that are possible with a computer, they permit the user to automatically retrieve information. The information is displayed in page size, usually on a viewing screen, or reproduced on a film or paper copy. However, data on the film cannot be moved from one location to another, nor rearranged or changed. (For further information, see chapter III.)

Microfilm Chip, Automated. This equipment has about the same capabilities as the system described immediately above. The use of the chips, however, does make it possible to insert and delete individual pages. (For further information, see chapter III.)

Aperture Card. (EAM punched card-microfilm). Systems of this type make it possible to mechanically sort, select, display, and copy printed or graphic information appearing on the film images displayed on the cards. However, as in the case of microfilm chip automated systems, the equipment is not well suited to personal searching by individual users. (For further information, see chapter III.)

Microform-Computer Combinations. Various types of microform equipment can be linked either directly or indirectly to a computer so that the computer can be used to conduct the searches and the microform device used to store and display the information or documents the user is seeking. (For further information, see chapter III.)

VI. HOW TO DECIDE IF A NEW SYSTEM IS NEEDED

The Preliminary Survey

This handbook gives considerable attention to finding the best system for storing and retrieving information. There will always be situations where the best system is the same system used in the past. Other situations will warrant the use of modern information retrieval methods and equipment.

Sometimes information retrieval studies are pursued for weeks or months, or a new system is installed, only to discover that a conventional system is all that is needed. The first question, therefore, that needs to be answered—and rather quickly—is "When do I use the old and when do I use the new?" This chapter describes a step-by-step procedure for making a preliminary survey to answer that question. It will help in deciding when conventional methods should be used and when it is worthwhile to spend the time and effort to make a detailed study of the possibilities of modern information retrieval methods and equipment.

Where to Look

The preliminary survey should not be limited to the major files, the library, or collections of reference materials. Rather, you should look anywhere there is a collection of information stashed away, regardless of the form in which it is stored. In this handbook, these files or other collections are referred to as "information facilities." Certainly, the size and frequency of use of the information facility are considerations, but they are less likely to rule out any system than they are to affect the type of system needed when weighed on the cost-benefits scale. Small units can sometimes justify relatively inexpensive and yet modern information retrieval systems. This is particularly true where there are many small information facilities containing information all or a substantial portion of which is the same.

For further clarification of the wide potential, consider any of the following situations:

Case-type records used to correlate or compare data relating to individual persons, places, or things, for such purposes as personnel selection and placement, selection of contractors for bidding, selection of equipment, and conducting special analyses.

Case-type records used for looking up and extracting discrete data such as names, addresses, amounts, dates, and other data needed for such purposes as answering correspondence, processing applications, and preparing reports.

Subject files and indexes relating to written text and used for obtaining any information that might aid in handling a current task or problem in connection with such activities as legal work, research, preparation of instructions, and management planning.

Reference collections containing such items as publications, technical reports, procedural manuals, directories, catalogs, and statistics used in day-to-day operations or research.

Files of graphic or pictorial material such as maps, photographs, slides, and engineering or architectural drawings in situations where the users are trying to find items having set characteristics or attributes.

Examining User Needs

Looking at all information facilities, of whatever description, is a practical and solid starting point. It is, however, at least equally important to examine the needs of the people who use the information.

Why is it important to look at both the infor-

SAMPLE PRELIMINARY SURVEY FORM FOR INFORMATION FACILITIES

Information Retrieval Preliminary Survey INFORMATION FACILITY	INSTRUCTIONS - Prepare one of these Reconnaissance Data Sheets for each file station, record collection, index file or other information facility at the installation being surveyed. Where reference is made to user manhours, specify those spent by employees of the facility as well as any spent at the facility by personnel from other organizational units.	ANALYSTS NAME DATE	
ORGANIZATION AND FACILITY NAME AND ADDRESS OF ORGANIZATION OF JURISDICTION	INFORMATION CONTAINED IN RECORDS	RECORD PHYSICAL FORM	
	TOTAL NO. (*Net*) EMPLOYEES ASSIGNED TO OPERATE AND MAINTAIN FACILITY	BLDG & ROOM NO.	PHONE NO.
TITLE OF INFORMATION FACILITY	CURRENT VOLUME (*No. of pages, if written info; No. of characters if precise data; No. of items, if graphics*)	ANNUAL ADDITIONS	USEFUL LIFE OF INFORMATION

USAGE DATA - To estimate manhours spent at this information facility in looking up, searching, extracting, or correlating information or data - - - - including manhours spent by personnel assigned to the facility as well as time spent at the facility by personnel from other organizational units.

PRIMARY USERS (*Organization and Unit*)	JOB TITLE	TOTAL ANNUAL MANHOURS	TYPE OF RETRIEVAL ACTION

DESCRIBE TYPICAL LOOKUP, SEARCH, DATA EXTRACTION OR CORRELATION ACTIONS

DESCRIBE ANY INADEQUACIES, PROBLEMS, OR LIMITATIONS OF THIS INFORMATION FACILITY

REMARKS

Figure 34

mation facilities and the users' needs? Why is it not sufficient to stop with a look at the demands upon and limitations of the information facilities themselves? There are many reasons, but the following are particularly significant:

- Data gathered at the information facility or from the users alone would be incomplete and misleading; whereas gathering information from both serves to supplement and cross check the information furnished by the other.

- Personnel operating an information facility cannot always describe or interpret user needs accurately.

Users' statements must be weighed in the light of actual information facility experience:

- If the information facility receives moderate or heavy use, the users probably have a real need for information—perhaps for even more than they are now getting.

- If the facility receives only light use, the probability of an urgent users' need is suspect unless the facility is not readily accessible nor operated properly.

Fact-Gathering Forms

The person conducting the preliminary survey should, if possible, personnally collect the data relating to the information facilities and users' needs, in which case the data could be recorded directly on decision tables similar to those shown in figures 36 and 37. If, however, the information users and the personnel operating the information facilities will be requested to supply the data themselves, the use of forms similar to those shown in figures 34 and 35 is suggested.

Information Retrieval Preliminary Survey—Information Facility (Fig. 34). This form may be used for collecting data about the various file stations, manual or machine record files, publications, and any other collections of typed, handwritten, printed, or graphic material. The data appearing on these forms, together with the personal knowledge of the individuals who completed them, will later serve as the basis for preparing information facility decision tables.

Information Retrieval Preliminary Survey—User Needs (Fig. 35). This second form may be used to obtain a sampling of how much time the users are now spending in looking up, searching, extracting, or correlating information or data, and to identify any inadequacies, problems, or limitations of the present sources or methods. These completed forms will also be used later for preparation of decision tables.

Decision Tables

Two decision tables have been prepared to help show what conclusions may be reasonably drawn from any set of facts gathered. These tables require the answering of various "yes" or "no" questions about the facts. The patterns shown by the "yes" and "no" answers lead to certain predetermined conclusions shown on the forms. One table is for analyzing facts gathered about the information facility and the other relates to facts about users' needs. Blank copies of these two decision tables are included as Appendix "D." Figures 36 and 37 provide filled-in examples of the two tables.

Evaluating Information Retrieval System Potential—Information Facility. (Fig. 36).

This form contains spaces for entries of certain identification and usage data at the top. Then, under "Evaluation Factors," a "Y" or "N" should be entered under the "Yes" or "No" column for each factor, depending upon your findings. The resulting yes-or-no pattern in this column is the same as one of the columns under "Key." It is this pattern that identifies the conclusion appropriate for the particular set of facts being analyzed. The "yes" and "no" answers might be thought of as "votes" for or against a modern information retrieval system (except for No. 5 evaluation factor, which is reversed). But it is not merely a matter of counting up affirmative and negative answers, since some evaluation factors carry more weight than others. It is the exception rather than the rule that the decision for or against would be based on just one of these factors.

SAMPLE PRELIMINARY SURVEY FORM FOR USER NEEDS

Information Retrieval Preliminary Survey USER NEEDS	INSTRUCTIONS: Prepare one or more of these Reconnaissance Data Sheets, as needed, for each of the broad, similar types of information or data needed by the installations being surveyed. *EXAMPLES: Personnel data, legal precedents, accounting data, research information, etc.*	ANALYST'S NAME DATE
BROAD TYPE OF INFORMATION OR DATA (Short title)	BRIEF DESCRIPTION OF THE INFORMATION OR DATA	

USAGE DATA – To identify individual user groups and the manhours they spend in looking up, searching, extracting, or correlating this information or data.

ORGANIZATIONAL UNIT	USER'S JOB TITLE *(Exclude personnel assigned to operate information facilities)*	NUMBER OF USERS	USER'S PHYSICAL LOCATION	ANNUAL MANHOURS *(All users)*	PRIMARY SOURCES OF THIS INFORMATION *(Name and location)*

DESCRIBE TYPICAL LOOK UP, SEARCH, DATA EXTRACTION, OR CORRELATION ACTIONS

DESCRIBE ANY INADEQUACIES, PROBLEMS, OR LIMITATIONS INVOLVED IN THE PRESENT SOURCES OR METHODS EMPLOYED

REMARKS

Figure 35

These are some of the basic concepts involved in the following "Evaluation Factors":

Factor 1: "Annual additions equal or exceed." Modern information retrieval systems are normally designed to handle fairly large collections of information or data. The addition of 25,000 pages or 2,500 individual graphic items annually or the maintenance of one million characters of data that are constantly being updated may be considered the minimum volume requirement for a positive vote for modern information retrieval methods. It is possible to have less volume and still find some need for an information retrieval system, but the probabilities are less likely. A "no" vote, therefore, does not necessarily rule out the potential need for an information retrieval system.

Factor 2: "Information will be in continuous use for over 5 years and one man-year or more is being used for looking up, searching, extracting, or correlating information or data at this facility." Because information retrieval systems always create new and often considerable expense, particularly in the input phase, they are ordinarily not used for information or data of short term value. And unless coupled with at least 1 man-year of work in searching, etc., there may not be enough potential manpower savings to offset the cost of an information retrieval system. A "yes" answer here is another vote for information retrieval, but by no means a justification in itself.

Factor 3: "Information will be in continuous use for less than 5 years and two man-years or more are being used for looking up, searching, extracting, or correlating information or data at this facility." The extra expense of an information retrieval system might be justified even though the information or data were of shorter use value if there is a potential for saving two or more man-years of searching time. Evaluation factors 2 and 3 are mutually exclusive—in a given situation only one could apply. Also, of course, in some instances neither may apply. Also note, as explained in the second sentence under "Instructions" at the bottom of the form, that the man-hours include both those of the personnel assigned to operate the facility as well as to others who come to conduct searches at the facility.

Factor 4: "Time presently required for looking up, searching, etc., information or data at this facility is mainly attributable to limitations of conventional methods." A "yes" vote is used here only when it can be determined that the reasons it takes so much time to retrieve information are due to the inherent limitations of conventional methods, and that it should be possible to reduce retrieval man-hours by installing a modern information retrieval system.

The fact that extensive man-hours are being spent to obtain information need not mean that the conventional system is inefficient. It may simply be due to the heavy workload. (In some situations a conventional system can retrieve information faster and cheaper than a modern information retrieval system.)

To evaluate this factor properly, one must therefore clearly understand the inherent advantages and disadvantages or limitations of both conventional and nonconventional methods.

Factor 5: "The information maintained at this facility could be readily obtained from other source(s)." Be sure to note that a "yes" vote here is a vote against a modern information retrieval system. This factor is included in the decision table because other places where the same information is available are sometimes overlooked. Modern transmission methods and duplicating services may make it more practical to use another source instead of maintaining a duplicate facility. By pooling the resources used to maintain the duplicate or complementary information facilities, it may also be possible to install a modern information retrieval system.

There follows explanations for the five conclusions depicted in figure 36.

Conclusion A: "A modern information retrieval system seems a likely possibility." This means only that from your observation at the present time, you can conclude that there is a definite possibility it may be profitable to install a modern information system.

Conclusion B: "Likely that present or improved conventional methods will suffice." This means that you have eliminated any reasonable doubt as to the need for a modern information retrieval system.

SAMPLE FORM FOR EVALUATING INFORMATION RETRIEVAL SYSTEM POTENTIAL—INFORMATION FACILITY

Evaluating Information Retrieval System Potential

INFORMATION FACILITY

EVALUATOR'S NAME: GEORGE ADAMS

DATE: 9 - 3 - XX

ORGANIZATION AND FACILITY

NAME AND ADDRESS OF ORGANIZATION OF JURISDICTION:
Administrative Div.
Office of Gneral Counsel
Denver Regional Office

TYPE OF RECORDS: [✓] FOLDERS [ ] CARDS [ ] OTHER (Specify)

CONTENTS OF RECORDS: Claims applications, legal opinions and decisions, court records, and corres.

TITLE OF INFORMATION FACILITY: Central Legal Files

NO. (Net) OF EMPLOYEES AT FACILITY: 3

BUILDING AND ROOM NUMBER: Rm 600 Admin. Bldg

PHONE NO: X39290

USAGE DATA (Estimated manhours spent annually in looking up, searching, extracting or correlating information or data at this facility)

PRIMARY USERS (Organization & Unit)	JOB TITLE	ANNUAL MANHOURS	PRIMARY USERS (Organization & Unit)	JOB TITLE	ANNUAL MANHOURS
Claims Div	attorney	1000	Gen. Law Div.	Law clerk	6000
Litigation Div.	attorney	1500	Law Library	Librarian	2000

EVALUATION FACTORS	YES OR NO	KEY								
		✓								
1. Annual Additions Equal or Exceed: (Circle applicable letter, if any) a. 25,000 pages, if system is used mainly for storage of written information. b. 1,000,000 characters, if system is used for storage of precise data such as names, numbers, etc. c. 2,500 individual items, if system is used mainly for storage of graphic, pictorial, or other matter not covered above (Explain in remarks).	Y	Y	Y	Y	Y	N	N	N	N	–
2. Information will be in continuous use for over 5 years and one man-year or more is being used for looking up, searching, extracting, or correlating information or data at this facility.	Y	Y	N	N	N	Y	Y	N	N	–
3. Information will be in continuous use for less than 5 years and two man-years or more are being used for looking up, searching, extracting, or correlating information or data at this facility.	N	N	Y	Y	N	N	N	Y	Y	N	–
4. Time presently required for looking up, searching, etc., information or data at this facility is mainly attributable to limitations of conventional methods.	Y	–	Y	N	–	Y	N	Y	N	–	–
5. The information maintained at this facility could be readily obtained from other source(s) (Specify sources and locations under remarks).	N	N	N	N	N	N	N	N	N	N	Y
CONCLUSIONS										
A. Modern information retrieval seems a likely possibility.		(X)	X							
B. Likely that present or improved conventional methods will suffice.				X	X		X		X	X	
C. Likely that present or improved conventional methods will suffice; HOWEVER, also consider modern information retrieval systems (particularly those which use inexpensive equipment.)						X		X			
D. Consider discontinuance of either this or other duplicate facility(ies), and if duplication is widespread, also consider possibility of a central information service or facility.											X
E. Other (Specify and explain - use remarks if additional space is required),											

REMARKS

1

INSTRUCTIONS - Prepare one of these Decision Tables for each file station, record collection, index file or other information facility at the installation being surveyed. Where reference is made to user manhours, specify those spent by employees of the facility as well as any spent at the facility by personnel from other organizational units. Answer "YES" or "NO" in the appropriate column opposite the Evaluation Factors to indicate the existing situation.

Compare your overall findings with those in the columns under the KEY. *(A dash indicates that it makes no difference whether the answer to that evaluation factor is Yes or No.)* When you find a column that duplicates your answers, place a check mar at the top of the column *(preferably with a colored pencil)*. Follow the appropriate column down into the Conclusions column and circle the appropriate X.

Figure 36

Conclusion C: "Likely that improved conventional methods will suffice; however, we should also consider modern information retrieval systems." This represents a "gray" area situation that you will probably not want to eliminate at this time.

Conclusion D: "Consider discontinuance of either this or other duplicate facility(ies), and if duplication is widespread, we should also consider the possibility of a central information service or facility." This is self-explanatory.

Conclusion E: "Other." This permits the person making the study to provide an alternate conclusion or to take exception to what would have been the normal conclusion due to factors not covered in the decision table: for example, if it were found that a major change in the functions, workload, or organizational structure were imminent.

Important: Note that the block at the top of figure 36, titled "No. (net) of Employees at Facility," refers to the net number of people (or man-hours) required for operating the facility, even though some situations may require only a small portion of the total staff for searching the files, the remainder being used to enter information into the system and keep it in proper condition. (This item should not be confused with the man-hour figures called for in evaluation factors 2 and 3.)

Evaluating Information Retrieval System Potential—User Needs (Fig. 37). This form is used and analyzed in the same manner as the information facility form in figure 36. These are the basic concepts involved in its evaluation factors.

Factor 1: "5 percent or more of users' total man-hours (minimum 1 man-year) are being spent in looking up, searching, extracting, or correlating information or data." The probability is that a modern information retrieval system will not be considered unless it can be justified economically. Hence, the more time that users spend in trying to get the information needed, the greater the possibility of saving their time and offsetting the cost of information retrieval systems. If the users spend less than 5 percent of their time in such efforts, it is unlikely that information retrieval can recover enough of the users' time to pay for the system.

Factor 2: "Current information facilities are inadequate for one or more of the following reasons." These represent disadvantages or deficiencies of conventional systems from the viewpoint of the users. Often these problems can be overcome through application of modern information retrieval methods. Factor 2 should be answered "yes" only when the problem is inherent in the conventional system employed, not when it is due to faulty design or operation. A "yes" vote here is therefore a vote for a modern information retrieval system.

Factor 3: "Much faster retrieval speed is needed than could ever be achieved under present or any other conventional method." If there is an overriding need for retrieval speed, there may be justification for a modern information retrieval system. This factor may be important enough to overrule negative responses to the other factors. Situations of this type often exist in intelligence work, defense systems, and sometimes in office areas, too.

Factor 4: "Time presently spent in searching, extracting, or correlating information or data is mainly attributable to limitations of conventional methods." The remarks for evaluation factor 4 for the information facility decision table also apply here. Further, a double check from the viewpoint of the user is necessary to make certain that the conventional system and equipment are the problem, rather than something else; for example, man-hours spent reading and examining documents after they have been retrieved, which is a common practice in some professions regardless of the retrieval system used. Therefore, to evaluate this factor properly the analyst needs to investigate present practices and procedures.

The explanations of the conclusions for this table are the same as the explanation offered for the table on information facilities, except for the omission of conclusion D, "Consider discontinuance of either this or other duplicate facility." This form also has an "Inconvenient Features" section at the bottom that is not part of the decision table itself but is supplementary in nature and is included for the following reasons:

- To make sure that the person making the study does not confuse mere inconvenience with inadequacy and thereby erroneously mistake the former for the latter in evaluation factor 2.

SAMPLE FORM FOR EVALUATING INFORMATION RETRIEVAL SYSTEM POTENTIAL—USER NEEDS

Evaluating Information Retrieval System Potential
USER NEEDS

EVALUATOR'S NAME: HELEN DAVIS
DATE: 9-3-XX

BROAD TYPE OF INFORMATION: Personnel data (employee skills, education, experience, etc.)

ORGANIZATIONAL UNIT	USER'S JOB TITLES (Exclude personnel assigned to operate information facilities)	NUMBER	PHYSICAL LOCATION	ESTIMATED ANNUAL MANHOURS	PRIMARY SOURCES OF THIS INFORMATION
Personnel Div	Placement Offic.	3	Rm 708 Main	1500	Central files
" "	Manpower Sect.	6	Rm 723 "	2000	" "
Misc Supervisors	Misc.	25*	Misc locations	500	" "
Personnel Div	Training Off.	3	Rm 714 Main	800	" "

EVALUATION FACTORS	YES or NO	KEY						
		✓						
1. 5% or more of users* total man-hours (minimum 1 man-year) are being spent in looking up, searching, extracting, or correlating information or data. (*Users include all persons who personally do the looking up, searching, extracting or correlation, EXCEPT those assigned to operate the Information Facilities)	Y	-	Y	Y	Y	Y	N	N	N
2. Current information facilities are INADEQUATE for one or more of the following reasons: (Circle any that apply) A. Pertinent documents or information are regularly being missed or system produces too much non-relevant material or information. (B.) System can furnish documents, only, whereas users would like to receive only portions thereof or precise data. (C.) System cannot satisfy need for retrieving precise data and correlating it.	Y	-	Y	Y	N	N	Y	Y	N
3. Much faster retrieval speed is needed than could ever be achieved under present or any other conventional method.	N	Y	N	N	N	N	N	N	N
4. Time presently spent in looking up, searching, extracting, or correlating information or data is mainly attributable to limitations of conventional methods.	Y	-	Y	N	Y	N	Y	N	-
CONCLUSIONS									
A. Modern information retrieval system seems a likely possibility		X	(X)		X				
B. Likely that present or improved conventional methods will suffice.						X		X	X
C. Likely that present or improved conventional methods will suffice; HOWEVER, also consider modern information retrieval systems (Particularly those which use inexpensive tools)				X			X		
D. Other (Specify and explain)									

INCONVENIENT FEATURES (Features NOT necessarily attributable to limitations of conventional methods. CHECK ANY THAT APPLY.)	
	DIFFICULT TO OBTAIN ACCESS TO INFORMATION
	USERS PREFER TO SEARCH BUT FIND SYSTEM DIFFICULT TO UNDERSTAND OR USE
	USERS NOT ROUTINELY INFORMED OF NEW INFORMATION PERTAINING TO THEIR WORK
	OTHER (Specify and explain) Records too far removed from users ✓

REMARKS

Unknown factors are the possible losses in operating efficiency, program effectiveness and manpower management caused by the inadequacies & limitations of present system.

INSTRUCTIONS - Prepare as many of these Decision Tables as needed to collect data during the course of surveying individual user groups to estimate manhours spent in looking up, searching, extracting, or correlating information or data. Summarize your findings by preparing one Decision Table for each of the broad, similar types of information required at the installation being surveyed.
Enter "YES" or "NO" in the column opposite each of the Evaluation Factors to indicate existing conditions.
Compare your overall findings with those in the columns under "KEY" until you find a set that matches yours - place a check-mark at the top of that column (preferably with a colored pencil). Follow the selected column down to the "CONCLUSIONS" and circle the appropriate X.

Figure 37

- To serve as a ready reminder of future action that should be taken in addition to or independent of the installation of a retrieval system.

- To supplement the data in evaluation factors 2 and 4 in borderline situations by providing additional clues as to which system to select—a conventional or a modern information retrieval system.

All of the inconvenient features listed could probably be corrected by adjusting and improving the existing conventional system.

Summary

The forms shown in this chapter, like all the others appearing in this handbook, are offered as suggested working tools only, to be used by those conducting the information retrieval studies. They are designed to assist in data gathering, analysis, decisionmaking, and documentation of the study. The forms may be used in their present format or may be modified to suit the needs of individual agencies.

The decision tables are not intended to substitute for human judgment, but rather to aid in quickly identifying those situations where a modern information retrieval system may be justified. In order to apply them correctly, it is not only necessary to fully understand how they are to be used, as explained in this chapter, but also to have a comprehensive knowledge of the limitations and advantages of conventional systems. This was discussed briefly in chapter I; if, however, the person conducting the study has not had experience in designing and operating conventional filing and library systems, additional research in these areas should be conducted. It is recommended that the National Archives and Records Service (NARS) records management handbooks *Subject Filing, Files Operations,* and *File Stations* be reviewed, in any event, before undertaking the preliminary survey.

When conducting a preliminary survey, the study should begin with a look at the information facilities. However, the findings should be organized on the basis of the broad types of information needed rather than by organizational elements or file stations. The reason for this is that only in rare instances is any particular type of information of interest to only a single organizational element. Further, the information is often drawn from more than one source, and the same information is usually found in more than one information facility.

The person conducting the survey should identify the broad types of information needed by the users as early as possible and then relate to each type the user groups and the file stations that serve as the source of the information. The final decision as to whether there is a potential need for an information retrieval system thus takes into consideration the varying needs of individual user groups as well as problems incurred in the operation of the information facility.

The data gathered and the conclusions reached during the preliminary survey are not of course adequate for going ahead and installing a system. A large scale information retrieval study and system installation might typically consist of the following phases:

1. The preliminary survey
2. Determination of system requirements (the feasibility study)
3. Development of system concepts and preliminary system design
4. Determination of equipment requirements and selection of equipment
5. Development of detailed system design and recruitment of personnel
6. Acquisition of equipment and training of personnel
7. Implementation and testing of equipment and orientation of users
8. Evaluation of system performance, and periodic revision of system

This handbook does not attempt to cover all these phases, but instead concentrates on those matters peculiar to information retrieval or those presenting special problems in designing, installing, and operating an information retrieval system.

VII. HOW TO DETERMINE SYSTEM REQUIREMENTS

The data gathered during the preliminary survey is far too sketchy and unreliable to serve as the basis for determining system requirements. Consequently, it is necessary to go back to those areas where there was an apparent potential need for modern information retrieval methods and to obtain additional data in order to make a further, more detailed analysis.

Data Collection Techniques

The various techniques that might be used in collecting the data are described below. These techniques are intended to complement rather than duplicate each other, although some redundancy is always desirable in order to verify the findings. In a large scale study, all or most of these techniques might be employed. However, there will always be situations where the use of a certain technique is not permissible or perhaps not practical or necessary. The objective of the person conducting the study should be to obtain the needed data in the best way possible to assure its completeness and accuracy and at the same time to minimize interruptions in the work of the organization and the man-hours expended by users and others involved in the study.

Questionnaires. Questionnaires, although not an entirely reliable or satisfactory method for gathering data, can be quite helpful, particularly in the area of user needs. Considerable care and testing are needed in phrasing the questions and interpreting the results in order to avoid misleading or invalid conclusions.

Interviews. Some of the information will necessarily be obtained through interviews. Interviews are also a good way to gain an understanding of the working climate and the attitudes of the individuals and to follow up on questionnaires when necessary.

Observations. Some of the data needed to determine system requirements can be obtained through on-site observations. Data such as current file size, physical characteristics of the records, and the age of the current collection may be obtained in this manner. Personal observation is needed to ensure a good understanding of the situation and can also serve as a check against data obtained through questionnaires and interviews.

Reports. The questionnaires, interviews, and observations will not provide all the data needed. Data such as work volume, man-hours used, and record inventories may appear in existing reports. Consequently, the person conducting the study should look over the existing reports and utilize them whenever possible for obtaining needed data. Also, of course, data gathered in connection with the preliminary survey should be used in this phase of the study.

Work Counts. While work counts should be used sparingly, they may be essential for obtaining data not contained in any existing reports nor available through other sources. The work count may be needed to obtain or verify such data as input volume, man-hour requirements, time lag, number of searches, average searching time, and volume of information retrieved. The period of the work count will vary according to the particular situation, but normally it should not need to be longer than 30 days; such counts should employ sampling techniques rather than attempting to be a 100 percent check. In a large-scale study, consideration should be given to the use of mechanized techniques employing source data automation (SDA).

Suggested Questionnaires

Figures 38 and 39 are examples of questionnaires that might be employed for collecting information regarding user needs. Both the items in the form and question sections would more than likely have to be modified or rephrased to tailor the questionnaires to the particular organization under study.

User's Report, Information Requirements, General (Figure 38). A questionnaire such as this one might be used to obtain an overall picture of the user needs, work habits, preferences, information problems, and recommendations. Consequently, it tends to be complex and would probably require somewhat detailed explanations and examples of answers appropriate under various circumstances. A brief orientation, preferably through group discussion, is therefore needed in order for the users to properly understand the questionnaires and thus obtain worthwhile results. This orientation should be part of the "Users' Briefing" described later in this chapter—another good reason such a briefing is highly desirable.

A review of this questionnaire reveals that it is used to probe for facts that will have a vital impact on the design of any information system. The answer to question 8 may of necessity be only an estimate, unless there is sufficient time and need for requesting selected users to maintain a diary (daily log) for a specified period. Some of the questions are purposely redundant to a certain extent in that essentially the same information is occasionally asked for in different ways since some of the questions will not be fully understood by all the users.

It should be expected that the cooperation and quality in completing the questionnaires will range from very good to very poor—therefore, those conducting the study must be careful not to jump to conclusions but instead should give careful thought to the circumstances, environment, biases, and other factors that may have affected the way the questionnaires were completed.

Follow-up interviews are absolutely essential to effective use of the questionnaires. Interviews should be conducted for clarification of significant inconsistencies or errors and when a user obviously needs assistance in completing a questionnaire. Some questions, such as numbers 4, 7, and 9–16, may be designed to produce clues rather than complete answers and explanations; and therefore, these queries require follow-up discussions with individual users to obtain a full understanding of the situation and its possible impact on an information retrieval system.

User's Report, Work Unit Information Requirements (Figure 39). This second questionnaire might be used to obtain an across-the-board sampling of actual current information needs and user practices. It is designed to find out how the user goes about getting the information needed to complete a specific task, for example, processing a case, answering an inquiry, making a study, or writing a new procedure.

To decide how many tasks or work units are to be reported the following guidelines are suggested:

1. If the nature of the work is such that more than one task or work unit is completed each day, request the users to prepare five forms, i.e., one for the first task performed each day for the next five days after the briefing session.

2. If the individual task or work unit varies in length from one to five days, have the user report only on the first new task occurring after the briefing session.

3. If the individual tasks or work units are usually longer than five days, complete the form to show the related information activities for a one-week period or upon completion of the task, whichever occurs first.

It is also necessary to determine whether the questionnaire will be distributed to each user or only to certain ones. Whenever possible, most of the users should be asked to complete them. The three categories of information in this questionnaire are:

- Questions 1–5 seek information about the nature of the task, the end product, the character of the information needed, the way in which the user identified it, and where he went to get the information.

- Questions 6–8 cover information on how the user went about getting the needed information, the techniques used, and the man-hours involved.

SAMPLE QUESTIONNAIRE FOR USER'S REPORT ON GENERAL INFORMATION REQUIREMENTS

USER'S REPORT INFORMATION REQUIREMENTS, GENERAL	Complete each of the questions to the best of your knowledge. Enter N/A for questions, when not applicable.	DATE
1. NAME	JOB TITLE	
2. PRIMARY DUTIES OR RESPONSIBILITIES		3. LENGTH OF TIME IN THIS WORK

4. ANY SEASONAL OR OTHER PERIODIC PEAK PERIODS, WHEN INFORMATION NEEDS TEND TO INCREASE?

☐ NO ☐ YES *(identify)*

5. ARE THE TASKS AND INFORMATION OR DATA REQUIREMENTS DESCRIBED IN THE ATTACHED "USER'S REPORT, WORK UNIT INFORMATION REQUIREMENTS" TYPICAL?

☐ NO ☐ YES *(explain)*

6. HOW IS THE INFORMATION OR DATA GENERALLY USED IN COMPLETING YOUR WORK ASSIGNMENT?

☐ DIRECTLY INCORPORATED IN THE END PRODUCT ☐ OTHER

☐ DIRECTLY INCORPORATED IN THE END PRODUCT AS BACKGROUND INFORMATION

7. DO YOU PREFER TO DO YOUR OWN SEARCHING OR INFORMATION LOOK-UP, RATHER THAN HAVING SOMEONE OR A MACHINE DO IT FOR YOU?

☐ NO ☐ YES *(explain)*

8. CHECK APPROPRIATE BOXES AND COMPLETE ITEMS BELOW TO INDICATE NET TIME PERSONALLY SPENT IN OBTAINING INFORMATION AT AN INFORMATION FACILITY *(include time spent at your desk or work station and personal files)*

CHECK	ITEM	TITLE OF INFORMATION FACILITY OR SOURCE	LOCATION	MONTHLY ACTIVITY	
				NO. OF TIMES	TOTAL HOURS
	GENERAL SEARCHING FOR INFORMATION CONTAINED IN WRITTEN TEXT				
	RETRIEVAL OF SINGLE SENTENCES, PARAGRAPHS OR OTHER STATEMENTS CONTAINED IN WRITTEN TEXT				
	RETRIEVAL OF GRAPHIC OR PICTORIAL MATTER				
	LOOKING UP, COPYING, EXTRACTING, OR FURNISHING DISCRETE DATA *(such as names, numbers, dates, and quantitative or qualitative data)*				
	LOOKING UP CORRELATING, COMPARING, REARRANGING OR OTHERWISE MANIPULATING DISCRETE DATA				
	SCANNING PERIODICALS, REPORTS, AND OTHER MATERIAL TO KEEP ABREAST OF LATEST DEVELOPMENTS IN YOUR FIELD				

9. GENERALLY ARE PRESENT RETRIEVAL SPEEDS ADEQUATE FOR NEEDS?

☐ NO ☐ YES

TIME FACTORS:	REQUIRED	DESIRABLE

IF ANSWER ABOVE IS "NO," IDENTIFY INFORMATION THAT NEEDS RETRIEVAL TIME REDUCED

Figure 38

SAMPLE QUESTIONNAIRE FOR USER'S REPORT ON GENERAL INFORMATION REQUIREMENTS

10. ARE YOUR RETRIEVAL EFFORTS HAMPERED BY ANY OF THE FOLLOWING CONDITIONS? *(check appropriate item boxes)*

- ☐ FILING AND/OR INDEXING NOT KEPT CURRENT
- ☐ INSUFFICIENT INFORMATION OR DATA COVERAGE
- ☐ DIFFICULT TO GAIN PHYSICAL ACCESS TO INFORMATION OR DATA
- ☐ PHYSICAL FORM OR FORMAT OF MATERIAL IS INCONVENIENT
- ☐ FILE NOT READILY BROWSABLE
- ☐ SUBJECT CLASSIFICATION OR INDEXING SYSTEM DIFFICULT TO UNDERSTAND OR USE
- ☐ SUBJECT CLASSIFICATION OR INDEXING SYSTEM IS INEFFECTIVE
- ☐ MORE THAN A FEW DOCUMENTS CONTAINING RELEVANT INFORMATION OR DATA ARE BEING MISSED
- ☐ A GREAT DEAL OF THE INFORMATION OR DATA FOUND IS USELESS OR REDUNDANT
- ☐ ARRANGEMENT OR FILE SEQUENCE OF NON SUBJECT TYPE FILE IS NOT WELL SUITED TO YOUR NEEDS
- ☐ OTHER PROBLEMS

11. WHAT HAS BEEN THE EFFECT OF THE ABOVE PROBLEMS ON YOUR WORK AND THE EFFICIENCY OF THE OFFICE?

12. WHICH OF THE FOLLOWING CONDITIONS MOST CLOSELY CORRESPONDS TO THE SEARCH RESULTS YOU NEED WHEN RETRIEVING INFORMATION BY SUBJECT?

- ☐ RETRIEVAL OF ALL DOCUMENTS OR OTHER RECORDS THAT MIGHT BE CONSIDERED RELEVANT TO THE QUERY WITH THE POSSIBILITY THAT A CONSIDERABLE AMOUNT MAY PROVE TO BE NONRELEVANT
  (to avoid the possiblity of overlooking any relevant material)
- ☐ RETRIEVAL OF ONLY THOSE DOCUMENTS CONTAINING THE SPECIFIC INFORMATION OR DATA DESCRIBED IN THE QUERY WITH THE POSSIBILITY THAT DOCUMENTS OF VARYING DEGREES OF RELEVANCE MAY HAVE BEEN MISSED
  (to avoid retrieving more material than is really needed or can be readily used.)
- ☐ OTHER *(explain)*

13. WHICH ONE OF THE FOLLOWING DEGREES OF SPECIFICITY OR DEPTH OF SUBJECT MATTER BREAKDOWN FOR WRITTEN INFORMATION MOST CLOSELY CORRESPONDS TO YOUR NEEDS?

- ☐ LOW SPECIFICITY - BROADER THAN THE SUBJECT BREAKDOWN IN THE TABLE OF CONTENTS OF A TEXT BOOK OR MANUAL
- ☐ MODERATE SPECIFICITY - ROUGHLY EQUIVALENT TO THE SUBJECT BREAKDOWN IN THE TABLE OF CONTENTS OF A TEXT BOOK OR MANUAL
- ☐ HIGH SPECIFICITY - MORE SPECIFIC THAN THE TABLE OF CONTENTS OF A TEXT BOOK OR MANUAL

14. ARE THERE ANY PARTICULAR FUNCTIONS OR WORK PERFORMED BY YOUR ORGANIZATION, WHICH YOU BELIEVE COULD BE SUBSTANTIALLY IMPROVED OR PERFORMED AT LESS COST THROUGH THE APPLICATION OF MODERN INFORMATION RETRIEVAL TECHNIQUES?

☐ NO ☐ YES *(explain)*

15. ARE THERE ANY PARTICULAR TYPES OF ARTICLES, REPORTS, OR OTHER RECURRING MATERIAL ABOUT WHICH YOU NEED TO BE ROUTINELY INFORMED ABOUT IN ORDER TO KEEP ABREAST OF THE LATEST DEVELOPMENTS IN YOUR FIELD?

☐ NO ☐ YES *(describe)*

16. ARE THERE ANY INFORMATION OR DATA FILES NOW BEING MAINTAINED, WHICH MAY NOT BE NEEDED, IF AN INFORMATION RETRIEVAL SYSTEM WERE INSTALLED?

☐ NO ☐ YES *(identify)*

Figure 38 (Continued)

SAMPLE QUESTIONNAIRE FOR USER'S REPORT ON WORK UNIT INFORMATION REQUIREMENTS

USER'S REPORT
WORK UNIT INFORMATION REQUIREMENTS

Complete each of the questions to the best of your knowledge. Enter N/A for questions, when not applicable.

NAME	JOB TITLE	DATE

1 DESCRIBE THE TASK OR WORK UNIT THAT CREATED THIS NEED FOR INFORMATION

2 WHAT WAS THE OUTPUT?

3 PRIMARY CHARACTER OF INFORMATION SOUGHT:

- ☐ WRITTEN INFORMATION *(correspondence, directives, reports, and publications)*
- ☐ GRAPHIC OR PICTORIAL MATTER *(maps, drawings, and photographs)*
- ☐ QUANTATIVE QUALITATIVE AND OTHER DATA *(statistics, performance standards, costs, size, color, shape, etc.)*
- ☐ OTHER *(specify)*

4 WHAT PARTICULAR IDENTIFYING FEATURE, DESCRIPTIVE TERM, OR OTHER MEANS SERVED AS THE PRIMARY BASIS FOR IDENTIFYING THE INFORMATION OR DATA SOUGHT *(specify whether primarily documents or record title; name or numbers, names, titles or numbers or other people, places or things; subject topics; quantitative data, etc.)*.

5 IDENTIFY THE INFORMATION OR DATA FACILITIES OR SOURCES USED INCLUDING PERSONAL FILES

TITLE OF INFORMATION SOURCE	ORGANIZATIONAL AND PHYSICAL LOCATION	PHYSICAL FORM OF MATERIAL

6 WAS ASSISTANCE RECEIVED? ☐ NO ☐ YES — IF YES, ENTER NAMES OR TITLE AND ORGANIZATIONAL LOCATION

7 TYPE OF PERSONAL SEARCHING PERFORMED

		MANHOURS SPENT
	GENERAL SEARCHING FOR INFORMATION CONTAINED IN WRITTEN TEXT	
	LOOKING UP COPYING EXTRACTING OR FURNISHING DISCRETE DATA	
	LOOKING UP CORRELATING COMPARING REARRANGING OR OTHERWISE MANIPULATING DISCRETE DATA	
	OTHER *(specify)*	

8 METHODS AND MATERIALS USED IN PERSONAL SEARCHING:

- ☐ USED SUBJECT INDEX CARDS
- ☐ USED PRINTED INDEX AND OR TABLE OF CONTENTS
- ☐ BROWSED ENTIRE DOCUMENT FILE
- ☐ SCANNED CONTENTS OF FOLDERS OR OTHER DOCUMENTS ARRANGED BY SUBJECT TOPICS
- ☐ SCANNED CONTENTS OF FOLDERS OR OTHER DOCUMENTS ARRANGED BY NAMES OR NUMBERS OF PEOPLE, PLACES, OR THINGS
- ☐ OTHER *(specify)*

9 HOW LONG WAS IT FROM THE TIME YOU INITIATED ACTION TO GET THIS INFORMATION OR DATA UNTIL YOU OBTAINED IT?

10 HOW QUICKLY DID YOU ACTUALLY NEED THIS INFORMATION OR DATA?

11 TOTAL MANHOURS YOU PERSONALLY SPENT ON COMPLETING THE TASK OR WORK UNIT INCLUDING THE TIME SPENT ON OBTAINING INFORMATION OF DATA?

HOURS	MINUTES

12 HOW SUCCESSFUL WERE YOU IN OBTAINING THE NEEDED INFORMATION OR DATA?

- ☐ OBTAINED ALL OR MOST OF IT
- ☐ IMPORTANT INFORMATION OR DATA APPEARS TO HAVE BEEN MISSED
- ☐ INFORMATION OR DATA WAS NEVER FOUND
- ☐ OTHER *(specify)*

13 NEED FOR THIS INFORMATION:

- ☐ CRITICAL
- ☐ SIGNIFICANT
- ☐ MARGINAL
- ☐ OTHER *(specify)*

Figure 39

- Questions 9–13 request information about the quality of the search results and the relative importance of the information search to the overall completion of the task.

As in the case of the earlier questionnaire, there will be instances where it may be necessary or desirable to interview individual users to obtain additional information.

Data Summarization Techniques

As explained in chapter VI, the study findings should be organized on the basis of the types of information needed and then related to the user groups and the information facilities that serve as the source of the information. A form similar to the system requirement worksheet (figure 40) may be used for this purpose. Such a form can serve not only as a convenient means for organizing the data but also as a checklist to assure that nothing of significance has been overlooked. One system requirement worksheet should be prepared for each of the broad types of information needed by the installation under study.

The sample system requirement worksheet is divided into four parts, as follows:

Part A—Input and Storage, page 1.

Part B—Retrieval and Presentation, page 2.

Part C—Resources, pages 3 and 4.

Part D—General Improvements Needed, page 4.

In conducting the study, of course, the output requirements for the system must be determined before it can be decided what information will have to be stored. Consequently the data for part B, retrieval and presentation, would have to be gathered first or perhaps simultaneously with that for part A, input and storage. While the form is largely self-explanatory, the following notes are offered to assist in its use.

Part A, Input and Storage. In examining input and storage requirements, the nature and volume of material that would have to be entered into must be known; therefore, this part reflects not only the current situation but future expectations as well.

Item 1, Physical characteristics. The physical characteristics of the input must be known since they have a direct effect on the type of equipment that can be used and personnel requirements.

Item 2, File size factors. Since some methods and equipment have optimum limits on the volume of material that can be stored or involve high storage costs, file size is always an important factor.

Item 3, Intellectual characteristics. Knowledge of the intellectual characteristics is needed since the more complex the intellectual requirements, the more sophisticated the system may have to be.

Item 4, Source factors. The source factors, like physical characteristics, directly affect ease of input and the type of storage equipment. For example, if the documents or data are produced in-house and could be received in computer magnetic tape form, the possibilities would be quite different from those where the producer is an outside organization and the information is available in printed form only.

Item 5, Change factors. If changes to the information entered into the system will be necessary, this fact must be known, since making the changes could be difficult and expensive if certain methods and equipment were to be employed.

Part B, Retrieval and Presentation. In this part are compiled the data needed to provide a comprehensive summary of user needs.

Item 1, Search activity factors. Types of retrieval actions and volume are important factors, since there are usually practical limitations in the workload that each equipment class can handle. The location of the users and their proximity to each other are also factors that might cause one method or type of equipment to be impractical and another to be ideally suited to the situation at hand.

Item 2, Search intellectual characteristics. If the users ask for documents or data by case name or number, the intellectual requirements imposed on

SAMPLE FORM FOR SYSTEM REQUIREMENTS WORKSHEET

SYSTEM REQUIREMENT WORKSHEET

TYPE OF INFORMATION	NAMES AND LOCATIONS OF INFORMATION FACILITIES INVOLVED

PRIMARY TYPE OF DOCUMENT OR RECORD:

☐ WRITTEN INFORMATION ☐ GRAPHIC OR PICTORIAL ☐ PRECISE DATA (*names, numbers, dates, etc.*) ☐ OTHER (*specify*)

PART A - INPUT AND STORAGE

1. PHYSICAL CHARACTERISTICS (*Document or record to be entered into the system*)

a. DOCUMENT SIZE AND FORM (*3x5" cards, 8x10½" sheets, 6x9" bound books, etc.*)	
b. AVERAGE LENGTH OF INPUT DOCUMENT OR RECORD SEGMENT (*number of page, if text; number of characters or lines, if data; etc*)	
c. LENGTH OF LONGEST INPUT DOCUMENT OR RECORD SEGMENT (*number of pages, if text, number of characters or lines, if data; etc*)	

2. FILE SIZE FACTORS (*of information or data to be stored*)

	NO. OF DOCUMENTS OR RECORD UNITS	TOTAL PAGES, CHARACTERS OR LINES
a. PRESENT QUANTITY		
b. ESTIMATED QUANTITY IN TWO YEARS		
c. ESTIMATED QUANTITY IN FIVE YEARS		
d. MONTHLY GROWTH		

e. OBSOLESCENSE FACTOR (*period or event after which document or record is no longer needed*)

f. AGE OF CURRENT COLLECTION

3. INTELLECTUAL CHARACTERISTICS (*of information or data to be stored*)

a. ESTIMATED TOTAL NUMBER OF INDEXING TERMS, ATTRIBUTES, ETC. IN INDEX VOCABULARY — IF POSSIBLE, ESTIMATE AVERAGE NO. ASIGNED EACH DOCUMENT OR RECORD

b. STABILITY OF FIELD: ☐ STABLE ☐ SMALL AMOUNT OF CHANGE ☐ CONSTANTLY CHANGING

c. DEGREE OF COMPLEXITY: ☐ SIMPLE DATA ☐ COMPLEX DATA ☐ ORDINARY TEXT ☐ COMPLEX TEXT

d. SCOPE AND RANGE OF COVERAGE: ☐ NARROW ☐ MEDIUM ☐ BROAD

e. MEANS USED TO IDENTIFY AND/OR DESCRIBE DOCUMENT OR RECORDS PRIOR TO RECEIPT BY FACILITY (*title or number, author, abstract, case or record number, etc.*)

f. DEGREE OF REDUNDANCY OF INFORMATION OR DATA WITHIN THE FILE: ☐ LOW ☐ MODERATE ☐ HIGH

g. EXTENT OF INPUT EVALUATION OR SCREENING NEEDED

4. SOURCE FACTORS (*of information or data to be stored*)

a. ABILITY OF SOURCE TO FURNISH IDENTIFYING, DESCRIPTIVE, OR OTHER DATA IN MACHINEABLE FORM

b. EXTENT OF DUPLICATION

DUPLICATED DATA	HOW MUCH	LOCATION	TYPE OF SYSTEM	COMPATIBILITY WITH SYSTEM UNDER CONSIDERATION

c. PRIMARY PRODUCERS OF DOCUMENTS OR DATA

IDENTITY	LOCATION

5. CHANGE FACTORS

a. EXTENT TO WHICH INDIVIDUAL DOCUMENTS, THEIR DESCRIPTIONS, OR IDENTIFYING TERMS, AND/OR DATA TO BE STORED IN THE SYSTEM WILL HAVE TO BE CHANGED, UPDATED, ADDED TO, OR DELETED

Figure 40

SAMPLE FORM FOR SYSTEM REQUIREMENTS WORKSHEET

PART B - RETRIEVAL AND PRESENTATION

1. SEARCH ACTIVITY FACTORS

a. VOLUME — MONTHLY RETRIEVAL ACTIVITY

NO. OF USERS	ORGANIZATION AND LOCATION	TYPE OF ACTIONS	NO.	NO. OF DOCUMENTS, PAGES OR CHARACTERS RETRIEVED

b. PHYSICAL DISPERSION OF USERS (*in percentages*): SAME FLOOR, SAME BUILDING ____ %; SAME BUILDING ____ %; SAME BLDG. COMPLEX ____ %

c. SEARCH ACTIVITY PERIODIC FLUCTUATIONS - IF ANY (*describe*)

2. SEARCH INTELLECTUAL CHARACTERISTICS

a. WRITTEN INFORMATION

SEARCH SPECIFICITY: ☐ LOW ☐ MODERATE ☐ HIGH

EXTENT OF CORRELATION REQUIRED (*subject*): ☐ NONE ☐ LOW ☐ MODERATE ☐ HIGH

AVERAGE NO. OF INDEX TERMS USED PER SEARCH

b. DATA RETRIEVAL

COMPLEXITY: ☐ SIMPLE ☐ MODERATELY COMPLEX ☐ COMPLEX

EXTENT OF CORRELATION REQUIRED (*data*): ☐ NONE ☐ LOW ☐ MODERATE ☐ HIGH

AVERAGE NO. OF DATA ITEMS PER RETRIEVAL ACTION

c. IDENTIFYING FEATURES TO BE USED TO IDENTIFY INFORMATION (*name, number, document or record title, place or thing, attribute or other index term, etc.*)

d. OTHER (*specify*)

3. OUTPUT OR PRESENTATION, (*physical characteristics needed OR DESIRED*)

a. TYPE OF OUTPUT (*selected data or facts, document no.s., whole documents, selected portions, etc.*)

b. METHOD OF PRESENTATION OR DISPLAY (*manual display of document no.s. or index records, etc.*)

c. OTHER (*describe*)

4. SERVICE REQUIREMENTS

a. SPEED

MAXIMUM PERMISSIBLE PERIOD BETWEEN TIME INFORMATION, DATA, OR RECORD FIRST REQUESTED OR NEEDED AND TIME DELIVERED	MAXIMUM BATCHED PROCESSING OF REQUESTS PERMISSIBLE (*specify daily or weekly*)	IF A SEPARATE DOCUMENT REFERENCE INDEX FILE WERE TO BE USED, GIVE MAXIMUM PERMISSIBLE TIME FOR:	
		CONDUCTING A SEARCH	DELIVERY OF A DOCUMENT

b. CONVERTABILITY AND COMPATABILITY WITH OTHER AGENCY SYSTEMS & EQUIPMENT (*if essential - describe*)

c. ALTERNATE SEARCH METHODS (*if needed - explain*)

d. USER SELF SEARCHING: ☐ NONE ☐ DESIRABLE ☐ ESSENTIAL

e. BROWSABILITY (*describe special needs, if any*)

f. CURRENT AWARENESS (*if needed or desired, indicate type and frequency of service*)

g. OTHER SERVICE REQUIREMENTS (*describe*)

5. QUALITY REQUIREMENTS

a. ACCURACY OF EQUIPMENT OR DEVICE (*explain*)

b. EQUIPMENT RELIABILITY: ☐ NORMAL ☐ CRITICAL

c. PROTECTION AGAINST LOSS OF INFORMATION OR DATA STORED: ☐ ROUTINE PRECAUTIONS ☐ SPECIAL MEASURES NEEDED

d. WRITTEN INFORMATION RECALL - PRECISION RATIO: ☐ HIGH RECALL ☐ OTHER (*specify*) ☐ HIGH PRECISION

e. CURRENCY FACTOR (*specify how up to date the information must be*)

f. OTHER QUALITY REQUIREMENTS - IF ANY

Figure 40 (Continued)

SAMPLE FORM FOR SYSTEM REQUIREMENTS WORKSHEET

PART C - RESOURCES

CURRENT ANNUAL INFORMATION COST (Complete this part to summarize total current information retrieval resources and costs, which need to be taken into consideration, when developing the proposed system. Enter N/A for any item not applicable.)

1. PERSONNEL COSTS	a. SUPERVISION AND OPERATION OF INFORMATION FACILITIES	
	b. OTHER PERSONNEL COST INVOLVED IN STORING AND RETRIEVING THIS DATA	
	TOTAL ANNUAL PERSONNEL COSTS	
2. EQUIPMENT COSTS AND SERVICE CHARGES	a. RENTAL COSTS, IF ANY	
	b. DEPRECIATION	
	c. MAINTENANCE	
	d. OTHER EQUIPMENT COSTS	
	TOTAL ANNUAL EQUIPMENT COSTS	
3. SUPPLY COSTS	a.	
	b.	
	c.	
	TOTAL ANNUAL SUPPLY COSTS	
4. SPACE AND MISCELLANEOUS COSTS	a. SPACE	
	b. MISCELLANEOUS COSTS	
	TOTAL ANNUAL SPACE AND MISC. COSTS	
5.a. TOTAL GROSS CURRENT ANNUAL COSTS to be taken into consideration in development of the proposed system (*items 1 through 4*)		
5.b. LESS ANNUAL RESIDUAL COSTS, IF ANY FOR SERVICING EXISTING INFORMATION FACILITIES, WHICH MUST STILL BE REFERRED TO AFTER NEW SYSTEM IS INSTALLED		
6. ADJUSTED GROSS CURRENT ANNUAL EXPENDITURES FOR STORAGE AND RETRIEVAL OF INFORMATION, WHICH ARE AVAILABLE FOR APPLICATION TOWARD COST OF NEW SYSTEM		
7. ESTIMATED COSTS attributed to NOT being able to RETRIEVE and/or manipulate information WHEN NEEDED. (*describe*)		
8. VALUE OF USER MANHOURS, which could be saved, if modern information retrieval system was installed		
9. TOTAL Estimated net annual expenditures, which are available for application to cost of proposed information retrieval system (*add items 6 through 8*)		

CURRENT CAPABILITY

1. PERSONNEL	a. AVAILABILITY OF PERSONNEL TO DESIGN, INSTALL, AND PROVIDE TECHNICAL SUPERVISION OF AN INFORMATION RETRIEVAL SYSTEM			
	PRESENT JOB TITLES	GRADE	EDUCATION, TRAINING AND EXPERIENCE	PRESENT ASSIGNMENT

Figure 40 (Continued)

SAMPLE FORM FOR SYSTEM REQUIREMENTS WORKSHEET

1. PERSONNEL *(Continued)*

b. AVAILABILITY OF PERSONNEL TO OPERATE AN INFORMATION RETRIEVAL SYSTEM

PRESENT JOB TITLES	GRADE	EDUCATION, TRNG. AND EXP.	PRESENT ASSIGNMENT

2. MECHANIZED EQUIPMENT

a. AVAILABLE FULL TIME

QUANTITY	NAME AND MODEL	LOCATION	OWNED OR RENTED

b. AVAILABLE PART TIME *(use the same column heads above)*

			PERIOD/HOURS

3. SPECIAL INFORMATION AIDS OR TOOLS

DESCRIBE ANY SPECIAL AIDS OR TOOLS CURRENTLY USED IN STORING AND RETRIEVING INFORMATION, OR DATA, FOR EXAMPLE SUBJECT CLASSIFICATION OUTLINES; INDEX VOCABULARIES; DATA TABLES, ETC.

DESCRIPTION	OFFICE OF ORIGIN OR JURISDICTION

4. PHYSICAL AND OTHER FACILITIES

a. COMMUNICATION AND TRANSPORTATION *(mail, teletype, messenger service, conveyers, shuttle bus, etc.)*

DESCRIPTION	SPEED/FREQUENCY	COST

b. SPACE

QUANTITY AND LOCATION	AIR CONDITIONED	GOOD LIGHTING

c. ELECTRICAL POWER

CAPACITY	RESTRICTIONS, IF ANY

PART D GENERAL IMPROVEMENTS NEEDED (Prepare this form to indicate improvements, which should or could be made, regardless of whether or not a modern information retrieval system is to be installed. Identify the information facilities and groups involved in each instance.)

1 REDUCTION IN TIME LAG:
- ☐ INPUT - REDUCE DELAYS IN ENTERING INFORMATION
- ☐ OUTPUT - REDUCE DELAYS IN MAKING SEARCH OR DELIVERY OF ITEMS TO THE USERS

2 STAFFING:
- ☐ OBTAIN SPECIALISTS
- ☐ CONDUCT ADDITIONAL TRNG.
- ☐ ASSIGN ADDITIONAL PERSONNEL OR MANHOURS

3 ORGANIZATION AND CONTROL OF INFORMATION:
- ☐ INCREASE INFORMATION OR DATA COVERAGE
- ☐ ELIMINATE USELESS OR REDUNDANT INFORMATION
- ☐ UPDATE OR REVISE OBSOLETE SUBJECT FILE
- ☐ REARRANGE OR REORGANIZE CASE FILE
- ☐ CONSOLIDATE INFORMATION FACILITIES

4 USE:
- ☐ IMPROVE PROCEDURES AND FORMS FOR OBTAINING DATA
- ☐ RELOCATE FACILITY FOR BETTER ACCESSIBILITY
- ☐ USE ALTERNATE FACILITIES
- ☐ IMPROVE COMMUNICATIONS BETWEEN SYSTEM OPERATORS AND USERS
- ☐ ORIENT USERS IN USE OF PRESENT FACILITIES

5 OTHER *(describe)*

Figure 40 (Continued)

the system will be practically nil. If, on the other hand, the users ask for the documents on the basis of the subject topics or attributes, the method and equipment must have quite another intellectual capability. It is usually wasteful and more expensive to acquire equipment that has "intellectual" ability far exceeding that which is actually needed; or, in the opposite situation, it would be a grave mistake to install a system that fails to fully satisfy complex needs.

Item 3, Output or presentation, physical characteristics needed or desired. If the users must have the entire document, the demands on the system and equipment would be quite different than in a situation where they want precise data or desire to have the answers presented in special printed form or on a cathode ray tube (CRT).

Item 4, Service requirements. It is the throughput speed, rather than the speed at which equipment internal processing takes place, that is important to the user. Also, it is important that the person making the study be aware of any need for making the system compatible with other systems and equipment that may presently exist or are planned for the future.

One must be aware of any alternate search methods that may be needed because some of the user groups are at remote locations or do not need a system having as much retrieval capability as other groups. It is important to know the extent of user self-searching as opposed to searching by an intermediary, since this will be of concern in selecting the right method and equipment.

The person designing the system must also know whether it must be "browsable"—i.e., permits the operator to scan or skim through the system freely and at the same time to see the results of his search, rather than having to formulate precise questions and to wait a considerable period for the answers. Further, it is necessary to know whether a need exists for incorporating a current awareness or selective dissemination of information (SDI) capability in the system to automatically notify or forward information to employees when it has a bearing on their area of interest. If such a capability must be included, this would also have an effect on the method and equipment to be used.

Item 5, Quality requirements. If the system is to be used for conducting subject searches, it must be known whether the system should have high recall; that is, retrieval of all information that might be in any way pertinent, or high precision; i.e., retrieval of only that information that has a high degree of pertinency. (See chapter IX). If the system should operate somewhere between the two, this too must be known when the system is designed.

Part C, Resources. The purpose of this data is to determine the extent to which the costs, equipment needs, and personnel requirements for a new information system could be offset by expenditures, equipment, and personnel now being expended for storage and retrieval of information.

Current annual information costs. The person conducting the study needs to ascertain which of the current personnel and other costs for operating present information facilities and conducting searches could be applied to offset the costs for a modern information retrieval system. This should also take into consideration savings of users' time made possible through the introduction of modern information retrieval methods.

Current capability. It is necessary to know whether there are people available who would be capable of designing, installing, and technically supervising a modern information retrieval system; for if such talent is not present or could not possibly be obtained, it would be senseless to recommend installation of the system. Similarly, the person conducting the study must also take into consideration the qualifications of the personnel and the capability of any equipment that would be available, particularly if the system will be used for subject-type retrieval.

Part D, General Improvements Needed. The purpose of gathering this data is to isolate and identify weaknesses or failures in the present system that are not necessarily the fault of the type of system in use, but rather the way it is being managed and operated. The person conducting the study should review these conditions carefully since they too would affect the design of a new system and present their own particular problems, some of which may be overlooked or ignored on the assumption that the new system will automatically solve them.

If conditions such as inadequate staffing, work backlogs, user resistance, and poor utilization of existing facilities persist under the present system, the same thing could occur if a modern information retrieval system were to be installed. It is imperative, therefore, to consider all future plans and proposals in the light of any needless weaknesses or failures in the past in order to gain the ability to prevent the same thing from happening if the new system were to be adopted.

Final Review and Analysis of Findings

After all the system requirement worksheets have been completed, a review should be made of the manner in which the information needs have been grouped. The scope and content of each of the broad types of information should be scrutinized for the purpose of determining whether any adjustments need to be made; for example, consolidation of two or more broad types into an even broader type.

This final analysis and review is very important, since each of these broad types of information represents, in effect, a separate "information center" and will be individually considered in initially selecting the methods and equipment to be used.

Users' Briefings

It is during the data gathering and analysis phase that the users should be brought into the picture. This has several advantages:

- First, gaining their interest and understanding helps assure better cooperation and thus achieves better results from the questionnaires.

- Second, the potential users, through a newly acquired knowledge of information retrieval, may come up with potential applications and ideas that would otherwise have escaped the attention of those conducting the study.

- Third, establishing an early working partnership with the users goes a long way toward reducing problems that are likely to occur in the installation stage—particularly those involving lack of user acceptance and understanding of the new system.

Consequently, one or a series of briefings should be conducted for those users who the preliminary survey indicates have a potential need for modern information retrieval methods. The briefing should consist of the following three parts:

1. Background information.
2. An introduction to modern information retrieval theory and methods.
3. Illustrated presentations or demonstrations of information retrieval methods and equipment.

Use of General Analysis Techniques and Tools

The special tools and guidelines featured in this handbook are intended to implement and not to replace those normally used in conducting systems studies. They are designed to assist in tailoring studies to the particular factors and considerations involved in information storage and retrieval. It may still be necessary, for example, to use spread sheets and matrixes to compile and display the data collected.

It may also probably be necessary to prepare process, flow, work distribution, or operation charts—in other words, to employ many of the same techniques and tools commonly used in conducting any methods and procedures or systems study, particularly those pertinent to ADP or mechanization feasibility and application studies.

VIII. SELECTING THE RIGHT METHODS AND EQUIPMENT

Because there are such a variety of methods and equipment used in information retrieval, selecting the right one is never a simple or easy task. The process starts with the elimination of those methods and equipment classes that are clearly not suitable or practical. It ends with the comparison of the system requirements for the job at hand against the capability, characteristics, costs, and other features of the remaining classes.

Step 1. Selecting the Applicable Functional Category

The first task in the selection process, elimination of those methods and equipment classes not suitable or practical, may be accomplished by determining exactly what information retrieval function or functions the proposed system must perform. Once this is done, the person conducting the study needs to be concerned only with those methods and equipment classes which are normally used to perform that function or functions. To make the task easier, this chapter identifies the various methods and equipment classes according to four broad functional categories as follows:

Document Reference (DR) Systems. These systems are used primarily for subject-type searches to identify documents, persons, places, or things that are pertinent to the search questions. The user or person conducting the search is given the name or number of the document, person, place, or thing; and he then refers to the complete document or record to find out the details. Such systems are intended to quickly reduce a mountain of information to a manageable handful.

One example of a system performing the DR function is an electronic computer used in legal research to identify by the case name earlier court cases involving the same points of law and a situation similar to the one at hand. Another example is an optical coincidence system that is used to quickly identify those employees in the organization who possess the necessary qualification, characteristics, or attributes for a vacant position or special assignment.

Document Storage (DS) Systems. These systems are concerned mainly with the physical means for storing documents; the documents are arranged by some simple means such as titles or numbers. These systems cannot be used for conducting subject-type searches, but instead require that the user have a prior knowledge of the name, identifying number, machine address, etc. used to identify the desired document.

An example of a system performing the DS function is the microfiche system used by the research and development community for storage and distribution of technical reports. Another example is a video tape system used for storing applications and other important papers relating to housing loans.

Unified Reference-Storage (URS) Systems. These systems are, in effect, a combination of the first two functional categories. These systems are used mainly in situations where there is an urgent need to view the pertinent documents at the same time a subject-type search is being conducted. An example of a system performing the URS function is a microfilm system with photo-optical code used for storing technical correspondence and conducting searches on the basis of subject topics, contract numbers, names of equipment manufacturers, addressees, correspondence symbols, etc.

Data Fact Retrieval (DFR) Systems. These systems instead of merely referring the user to the name or number of the person, place, or thing, give the user the precise data or facts he is seeking. DFR systems are of two types—simple data lookup and complex data retrieval.

An example of a system performing the simple data lookup DFR function might be a mechanized roll microfilm system storing servicemen's allotment data and employing an odometer-type device to aid the user in quickly locating data relating to an individual serviceman. An example of a complex DFR retrieval system would be a computer system that maintains a large amount of data about each employee and then is used to compare, manipulate, select, and print data when conducting searches and preparing reports.

The decision chart depicted in figure 41 is intended as an aid in selecting the right (applicable) functional category, particularly for those who are conducting an information retrieval study for the first time.

Step 2. Selecting the Right Methods and Equipment

The second step consists of matching the system requirements as reflected in the system requirement worksheet against method and equipment capability, characteristics, cost, and other factors, as shown in the Nonconventional Methods and Equipment Guide, Appendix "A." Both this and the decision chart, figure 41, are designed to serve as only guides for quickly narrowing the wide, diverse fields of nonconventional methods and equipment to those few types that would normally be best suited to meet a particular set of system requirements and help make a final selection.

The nonconventional methods and equipment guide is organized in the same manner as the system requirements worksheet:

Part A—Input and Storage

Part B—Retrieval and Presentation

Part C—Resources.

The headings at the top of the columns on the guide refer to classes of equipment (not of any particular manufacturer). Part C, resources, must by necessity be completed by the person conducting the survey and is therefore separate.

After determining the appropriate functional category as explained above, it should be necessary to consider only those classes of methods and equipment marked "X" or "-X" for that functional category in the block immediately below the class title of the method or equipment. (However, there may be exceptional circumstances when one of the undesignated classes of methods and equipment will apply.) An "X" in the functional category block signifies that the particular method and equipment class is generally well suited for performing that function. A "-X", on the other hand, indicates that the method and equipment class might possibly be used to perform that function, but there may be limitations or other reasons it is less than ideally suited to many situations. (Descriptions of the various methods and equipment classes are included in chapters III, IV, and V.)

Because there will rarely be a situation where there is a perfect match between system requirements and equipment capabilities and characteristics, there usually will be a number of "tradeoffs" to analyze and weigh. In some instances, the nonconventional methods and equipment guide identifies capabilities in terms of "ideal," and the fact that the system requirements do not fall specifically within that range should not necessarily bar the use of that particular class, but instead may merely put it in the questionable category. Much of the success of any methods and equipment class, including those with a strong "yes," depends upon the ability of the system designer. A methods and equipment class that initially appears questionable may, through clever systems design, prove entirely satisfactory.

Some of the advantages of a particular equipment class will be offset or outweighed by its disadvantages, when its application to the situation at hand is considered. There may also be some situations where, due to an overriding need or other peculiarity, an equipment family other than that pinpointed by the tables may be more appropriate; however, the tables would still serve as a means for obtaining a summary of the capabilities, advantages, and disadvantages of other equipment.

After deciding which method and equipment would be best suited to meet the needs for each of the broad types of information needed by the users, the analyst should then examine the situation in terms of overall installation needs and

DECISION CHART FOR SELECTING THE RIGHT FUNCTIONAL CATEGORY

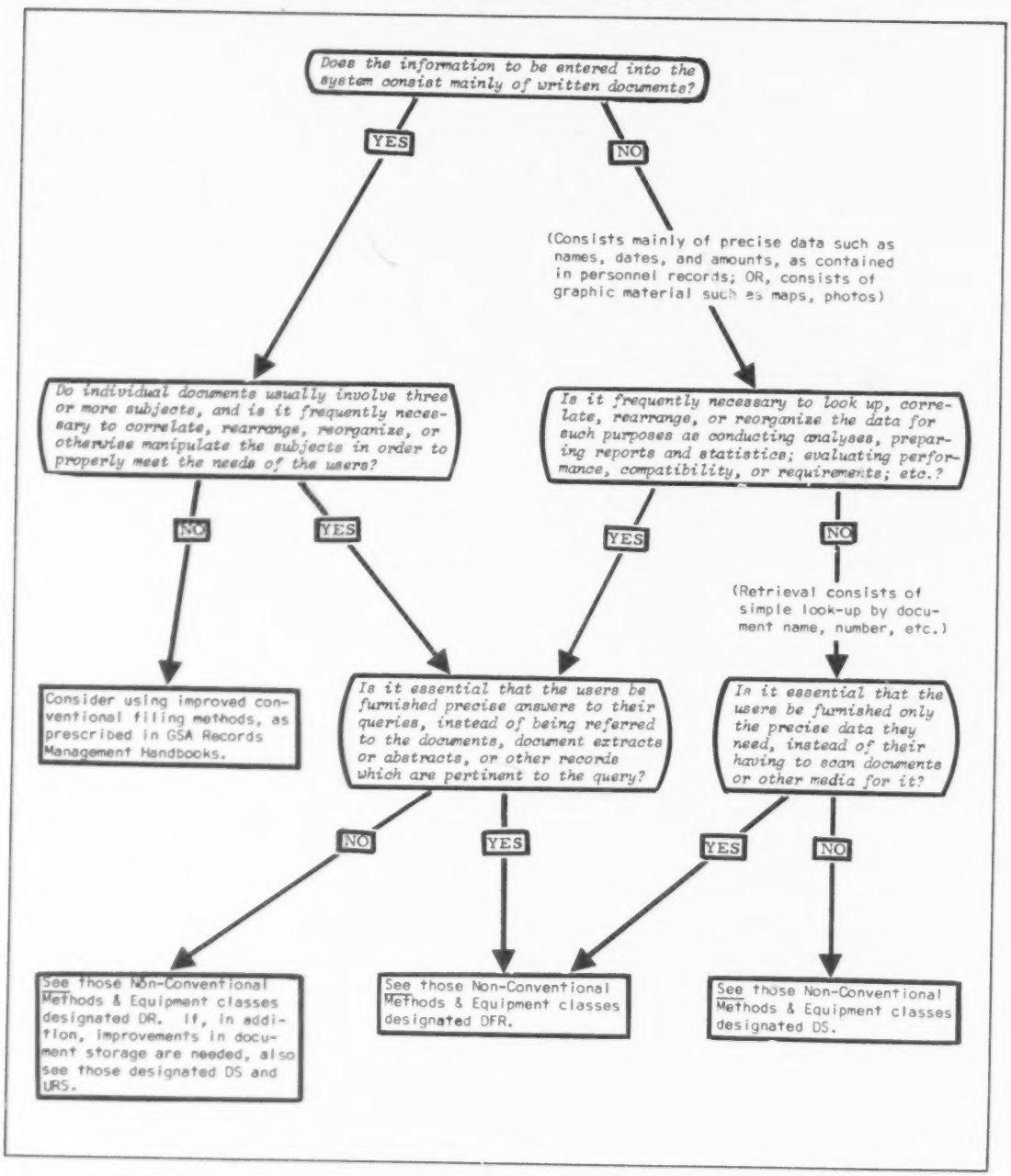

Does the information to be entered into the system consist mainly of written documents?
YES
NO
(Consists mainly of precise data such as names, dates, and amounts, as contained in personnel records; OR, consists of graphic material such as maps, photos)
Do individual documents usually involve three or more subjects, and is it frequently necessary to correlate, rearrange, reorganize, or otherwise manipulate the subjects in order to properly meet the needs of the users?
Is it frequently necessary to look up, correlate, rearrange, or reorganize the data for such purposes as conducting analyses, preparing reports and statistics; evaluating performance, compatibility, or requirements; etc.?
NO
YES
YES
NO
(Retrieval consists of simple look-up by document name, number, etc.)
Consider using improved conventional filing methods, as prescribed in GSA Records Management Handbooks.
Is it essential that the users be furnished precise answers to their queries, instead of being referred to the documents, document extracts or abstracts, or other records which are pertinent to the query?
Is it essential that the users be furnished only the precise data they need, instead of their having to scan documents or other media for it?
NO
YES
YES
NO
See those Non-Conventional Methods & Equipment classes designated DR. If, in addition, improvements in document storage are needed, also see those designated DS and URS.
See those Non-Conventional Methods & Equipment classes designated DFR.
See those Non-Conventional Methods & Equipment classes designated DS.

Figure 41

existing capability. The use of a reconciliation sheet similar to that shown in figure 42 is suggested for recording your findings and conclusions when matching individual system requirements against the capabilities, characteristics, and costs of the applicable methods and classes. It is suggested that the results be recorded as "yes" (Y), "no" (N) and "maybe" (?) in the blocks for parts A and B and also in the spaces for the overall conclusions. Part C, resources, of both the reconcilia-

SAMPLE RECONCILIATION SHEET FOR METHODS AND EQUIPMENT SELECTION

RECONCILIATION SHEET
SYSTEM REQUIREMENTS - METHODS AND EQUIPMENT

TYPE OF INFORMATION (USER NEEDS)

INSTRUCTIONS - Follow instructions in Figure 1 and page E-1, and the check blocks at top of pages E-12 through E-15, of "Non-Conventional Methods and Equipment Guide" to determine which method and equipment classes should be considered for storage and retrieval of the type of information (user needs) described above, and place an "X" above the same method and equipment classes on this form. Match each of the system requirements against the corresponding method and equipment capability for Parts A and B, part-by-part and item-by-item, and enter "Y" for yes, "N" for no, or "?" for undecided, in each of the columns marked "X" below, to indicate your conclusions. Delay comparison of method and equipment cost factors against current resources in Part C until the necessary information has been obtained from the manufacturers or suppliers for the classes marked "Y" or "?" in the OVER-ALL CONCLUSIONS for Parts A and B; however, be sure to complete the two availability factors at the bottom of Part C.

CHARACTERISTIC OR FACTOR		METHOD AND EQUIPMENT CLASSES																				
		CLUE-WORD EXTRACT CARD	PERMUTED INDEX	COLUMNAR CARD	DUAL DICTIONARY	EDGE NOTCHED CARD	OPTICAL COINCIDENCE	EAM PUNCHED CARD	MISC. CARD SELECTORS	MICROFILM JACKET	MICROFICHE	MICROFILM STRIP	MICROFILM ROLL MECHANIZED	MICROFILM ROLL PHOTO-OPTICAL	MICROFILM CHIP AUTOMATED	MICROFILM EAM PUNCH CARD	MICROFILM NOTCHED CARD	MICROFILM SUPERMINIATURE	VIDEO TAPE	COMPUTER ALL	COMPUTER MASS MEMORY
Part A INPUT AND STORAGE	PHYSICAL																					
	FILE SIZE																					
	INTELLECTUAL																					
	SOURCE																					
	CHANGE																					
Part B RETRIEVAL AND PRESENTATION	ACTIVITY																					
	INTELLECTUAL																					
	OUTPUT PHYSICAL																					
	SERVICE																					
	QUALITY																					
OVER-ALL CONCLUSIONS - Parts A & B																						
Part C RESOURCES (Current)	PERSONNEL COSTS																					
	OTHER COSTS																					
	PERSONNEL AVAILABILITY																					
	OTHER AVAILABILITY																					

Figure 42

tion sheet and the nonconventional methods and equipment guide should be completed only after the necessary information has been obtained from the manufacturers and suppliers or other sources for the classes marked "Y" or "?" on the "Overall Conclusions—Parts A and B" line of the reconciliation sheet.

When these analyses are concluded, the person conducting the study should be ready to submit his findings, conclusions, and recommendations to management. The best solution to the information problem in many instances lies in a combination of methods and equipment—some of which may be new and some of which may be old. It is not only prudent and practical to retain those features of the old system that the users prefer, but also of considerable help in gaining acceptance of the new system.

Other records management handbooks that should be helpful in conducting this phase of the study are *Information Retrieval Systems*, a description of 50 operating information retrieval systems in Government and private industry; *Microform Retrieval Equipment Guide*, which describes the capabilities, characteristics, and costs of microfilm readers and reader printers; and, the *Source Data Automation Equipment Guide*, which explains the various techniques and equipment for capturing or converting data to machine language for automated processing.

IX. DESIGNING A COORDINATE INDEX

Most modern information retrieval systems employ some form of coordinate indexing. This chapter is mainly concerned with designing coordinate indexes employing manual indexing and used for retrieval of documents on the basis of their subject matter content. However, most of the guidelines also apply to designing systems used for conducting searches to identify people, places, or things on the basis of their characteristics, features, or attributes. The objective of this chapter is to provide guidance on the subject of designing a coordinate index and highlight the main considerations.

Economics of Coordinate Indexes

Investment in Input Versus Output. In a conventional system where the high cost of retrieving documents is mainly attributable to the inherent problems and limitations of conventional methods and equipment, the chances are that too little is now being invested in the input. While increases in indexing (input) effort will have a substantial effect initially on reducing retrieval (output) costs, the return is diminishing. A point is ultimately reached where further savings in output is possible only at a great additional investment in input, thus making the total cost per retrieval action higher than for a conventional system.

The lowest overall cost in any given situation can be achieved only by a proper apportionment of investment between input and output. Because usually far more information is entered into a system than will ever be retrieved, it is often better to forego some of the refinements in input, such as sophisticated linguistical controls, in favor of doing a little more work at the output stage, such as screening the search results. Figure 43 illustrates a range of input-output cost relationships that a systems design should consider in determining the maximum cost-benefit for a particular system.

Input Costs. In coordinate indexing systems, the main input costs are labor. If the system employs manual indexing techniques and is used for retrieval of documents on the basis of subject topics, the input effort is largely intellectual—man-hour requirements for analyzing incoming documents, and assigning index terms. If the system employs automatic indexing techniques or is used for identifying people, places, or things on the basis of their characteristics, features, or attributes, the major costs are for clerks and machine operators—man-hours for entering the information into the system. In both instances, system design and application of source data automation (SDA) techniques play a vital role in controlling input costs.

Effort Versus Results. It is important to recognize that in information retrieval, the total effort put into the system is subject to the laws of diminishing returns. No matter how much effort is put into collecting, organizing, and processing the information, the system itself will never be able to satisfy all the users' needs. There will always be instances where it may be more practical to rely on special handling, for example, consulting experts or other information sources or services for assistance.

Steps in Developing a Coordinate Index

While the methods used in developing a coordinate index will vary in accordance with the time available, the complexity of the situation, and other factors, there are certain essential steps. The sequence of the steps may vary from that shown below, and it is usually desirable to undertake some of these steps simultaneously:

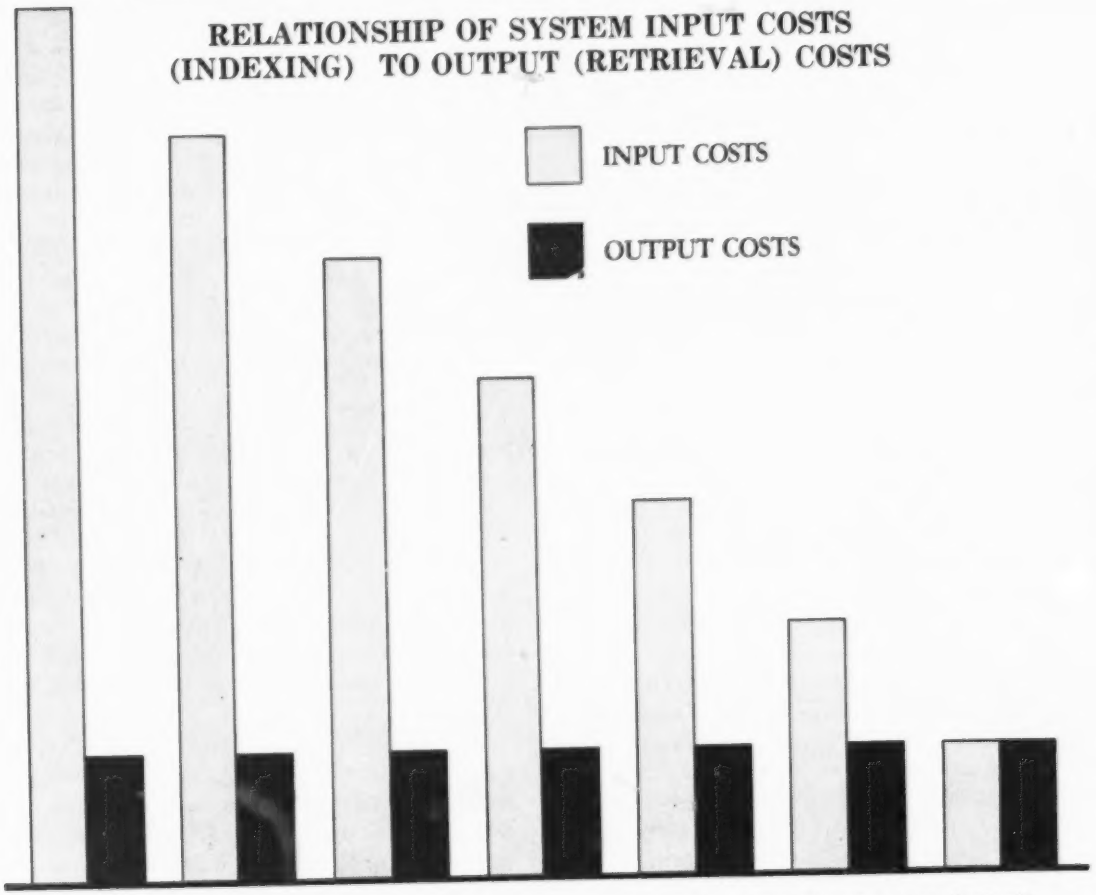
1. Review existing vocabularies.
2. Sample the documents.
3. Sample present searches.
4. Draft preliminary vocabulary.
5. Set up temporary index file.
6. Test and refine vocabulary.
7. Prepare the index manual.

Step 1: Review Existing Vocabularies. When developing a coordinate index one must be careful to select indexing terms on the basis of their significance in the subject matter field involved and their usefulness in conveying needed concepts. The review of existing vocabularies should include not only the formalized lists of descriptive terms, but also any other items that contain terms peculiar to the user group. Consequently, these sources of vocabulary material should be reviewed:

- Agency subject—classified outlines, subject indexes, or similar items.
- Organizational and functional charts and statements.
- Agency or installation annual reports and other publications describing the work of the organization.
- Laws, regulations, and directives.
- Index vocabularies in the same subject matter field developed by other Government agencies and private industry.

Step 2: Sample the Documents. A sampling should be made of the actual documents to be entered into the system in order to obtain a good idea of the range, scope, depth of coverage, and terminology used. If there are seasonal factors or other special circumstances, the selection method should be adjusted as necessary to obtain a representative sampling.

Step 3: Sample Present Searches. It is important to carefully study the present searches being made in order to obtain a good understanding of user language, preferences, and work habits.



Which of these relationships provides maximum cost-benefit at a particular installation?

Figure 43

The person developing the index should find out from the users themselves precisely what information they are seeking, why they want it, how they go about getting it, and how they use it in their work. Documents and other sources of information should be examined in order to determine how well they meet the users' needs both physically and intellectually. (Aids to be used in conducting this step are described in chapter VII.)

Step 4: Draft a Preliminary Index Vocabulary. Perhaps the ideal way to develop an index vocabulary might be to let it evolve slowly and naturally by selecting as indexing terms those words appearing in the titles, abstracts, and text of the documents. Then, through a continuous process of review, analysis, and refinement, one could produce the official vocabulary. However, this may take more time, both in terms of man-hours and elapsed time, than is available for vocabulary development.

Therefore, the system designer may prefer to draft a preliminary vocabulary of terms to bring a little order into the system at the beginning and to speed up the process of developing the coordinate index. In evaluating candidate terms consideration should be given to (1) their anticipated frequency in indexing and searching; (2) their relationships to other terms in the vocabulary; and (3) their acceptability as authentic terminology in the subject discipline concerned.

The preliminary vocabulary of indexing terms need be little more than a rough list of terms that previous reviews and samplings have demonstrated to be of importance. The list should be prepared in such a way as to permit insertion of terms at any point. A card file is probably the simplest way to do this. Definitions should be applied where needed, including instances where a term has more than one meaning. These include parenthetical qualifiers and, if more clarification is needed, a short definition called a "scope note." Where there are obvious synonyms the most suitable term should be selected for the index vocabulary, with a cross-reference to the synonymous term.

All vocabulary refinements and restraints should be limited to those that are needed to get the system started. Inclusion of any unnecessary refinements or restraints at this time will tend to interfere with the evolution of the vocabulary during the testing and refinement steps. However, the system designer should retain all the terms he has collected since it will be necessary to include many of them in the index manual for cross-reference purposes.

Keywords or descriptors? There are two basic types of indexing terms:

- Keywords, which consist of actual words appearing in the title, abstract, or text of the document.
- Descriptors, which consist of terms which have been developed for use in a particular system, and which by coincidence will also occasionally appear in the title, abstract, or text of a document.

Keywords provide a simple, rapid method of high specificity or deep indexing, since all the indexer has to do is underline significant words appearing in the title and the text. They also tend to preserve the actual words or "natural language" of the author, rather than subjecting his words to interpretation and conversion to standard terms by an indexer. Keyword systems, however, result in loss of information when applied to large collections of material or a wide variety of subjects, because of the author's inconsistencies, which might include such things as use of different words to mean the same thing, use of the same word to mean different things, and use of personal metaphors and jargon. The excessive number of keywords that are certain to accumulate and the scattering of similar terms because of the author's inconsistency usually cause such systems to eventually become unwieldy and difficult to search. Therefore, keyword systems are not generally recommended for manual indexing systems.

Descriptor vocabularies, if properly developed, controlled, and applied, are easier to search and more accurate than keyword systems. Descriptor systems make it possible to index concepts that may be missed when only the words and the document are used as indexing terms. Because fewer indexing terms per document are needed, the total

number of indexing terms is far less than in keyword systems. Descriptors reduce the size of the index file and thereby save index storage and equipment costs and searching time. However, development and control of the descriptor indexing vocabulary requires professional know-how and trained, skilled indexers. Poor design, inconsistencies, and errors can reduce retrieval accuracy or even nullify the advantages of descriptor systems over keywords. Further, indexing is more time consuming and tedious than keyword systems since it involves subject analysis, looking up indexing terms in a glossary or thesaurus, and making decisions about which descriptor to use.

Keywords and descriptors may be used to considerable advantage in the same indexing system. When this is done descriptors usually serve as the main, official index vocabulary. Then, when indexing documents, the indexers are permitted to supplement the assigned descriptors with any keywords that they have learned from experience might be particularly helpful in retrieving the document later. Such keywords as trade names, popular jargon, and coined terms can thus be added to the index description of the document without disturbing the operation of the basic descriptor system.

Hard language or soft? A "hard" language or vocabulary is one in which the indexing terms are straightforward, well defined, and readily understood. Such terms as physical characteristics, quantitative measures, and geographical locations would produce a "hard" vocabulary by their very nature. Unfortunately, much of the language contained in documents to be retrieved by subjects is vague, imprecise, inconsistent, and abstract. Because such "soft" language invariably creates serious problems in indexing and in searching, one of the primary objectives in the construction of the index vocabulary should be to convert "soft" language to a more precise "hard" vocabulary of indexing terms.

Hardening of the vocabulary is accomplished by: (1) Careful treatment of synonyms and near-synonyms by deciding which term will be used and then cross-referencing the others to it. (Near-synonyms refer to words that have different dictionary definitions but which are frequently used interchangeably; for example, "mechanized" and "automated."); (2) Avoiding the use of terms that are not meaningful or which are so vague as to defy precise definition; (3) Developing clear definitions; (4) Using common standard technical terms, if they exist, in preference to trade names, lay terms, and short-lived coined or popular terms; (5) Using root words; that is, using the simple form of a word to cover all of its variations, sometimes referred to as "confounding"; for example, the word "extend" might include "extension," "extensive," "extended," and "extending"; (6) Using the noun form for all indexing terms; for example, use "pouring" instead of the verb "pour"; and (7) Using the plural rather than the singular form, except when referring to specific processes, properties, and conditions.

Step 5: Set Up a Temporary Index File. The index file is the medium upon which are recorded the indexing terms and other descriptive data used to identify individual documents. Columnar cards, optical coincidence cards, and computer magnetic tapes are some examples. Usually the temporary index file is of the same type that will be used for the permanent index record; however, in smaller files at least, simple handposted columnar cards may be used. Special measures should be taken, to the extent possible, to facilitate changes, additions and deletions in the temporary file. Steps should also be taken that will later permit incorporating the temporary file into the permanent file without having to redo the work. One of the ways to accomplish this is to prepare and retain individual paper tapes or EAM punched cards for the document as it is entered into the system during this period.

Arrangement of the coordinate index file. There are two basic ways for arranging the index file: (1) by document title or number; or (2) by indexing term or term number. (See figure 44.)

Document or conventional file arrangements consist of one index card or individual machine record for each document or item being indexed. All indexing terms and other descriptive data for a particular document or item are usually posted to its one index record. The index file is arranged by document title or number. The conventional file arrangement preserves the indexing of each individual document or item as an integral unit that can be helpful in analyzing the index file and correcting or changing index postings.

COORDINATE INDEX FILE ARRANGEMENTS

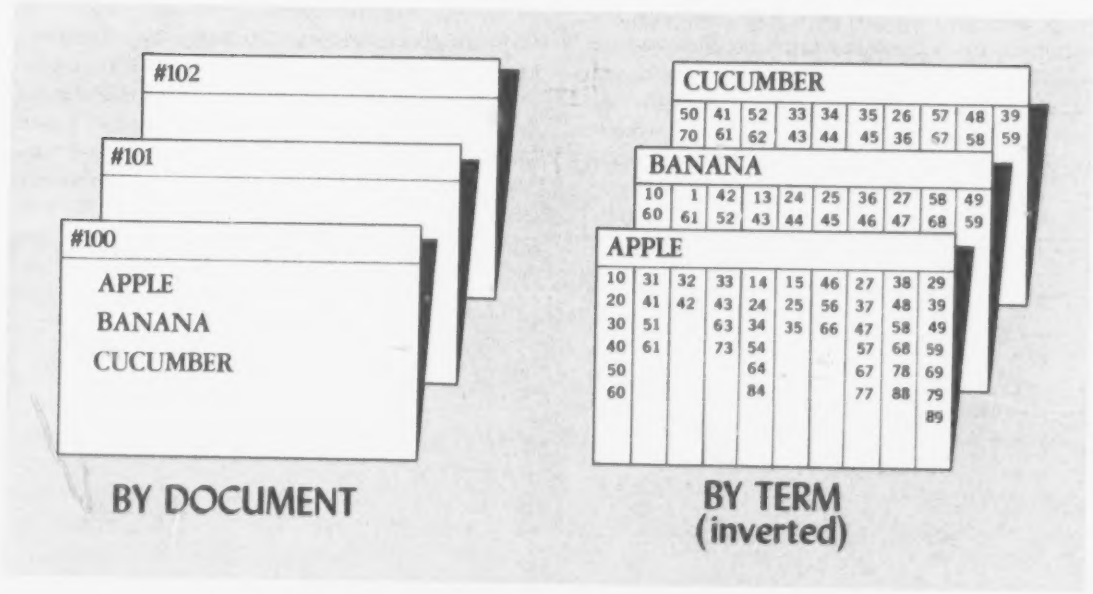


Figure 44

Conventional file arrangements also make it possible to have immediate knowledge of all the indexing terms assigned to the documents selected by the system during the search process, thus providing valuable clues as to their relevancy to the search question. However, such file arrangements require a large number of index records since there are usually several times as many documents or things to be indexed as there are indexing terms. For example, if there are 20,000 documents and 2,000 indexing terms, 20,000 index records would be needed. Conventionally arranged files require linear or serial searching of the file, which is usually more time consuming since every index record must be examined to determine if it contains the index terms used in the search question. For example, if only five indexing terms were used in the search, all 20,000 index records would still have to be examined.

Term or inverted file arrangements consist of one index card or individual machine record for each indexing term in the index vocabulary. The file is arranged by the indexing terms or term numbers. When the indexer has decided which indexing terms apply, the index records for those terms are selected and the document or item number is posted on each applicable index record. The inverted index file arrangement reduces to a minimum the number of individual index records that must be maintained. For example, if there are 20,000 documents of items and 2,000 indexing terms only 2,000 index records would be needed. Inverted file arrangements also greatly reduce the number of index records that must be examined, and thus also reduce the time required for the search process; for example, if five indexing terms were to be used in the search, only five index records would have to be examined. The major disadvantage of inverted file arrangements is that a search produces identifying numbers only, and it is therefore necessary to refer to another record to obtain descriptive information about the document and to determine its relevancy to the search question.

Term (inverted) and document (conventional) file arrangements are both sometimes used in the same system, particularly in those employing computers. The inverted file of indexing terms is maintained on-line to the computer to permit rapid searching of the entire file at one time. The search questions and the document numbers produced as a result of the search are then batched and periodically machine processed across a conventionally arranged magnetic tape index file con-

taining the complete bibliographic information for each document in the system, possibly including an abstract. Thus the user can be furnished a printout showing the results of the search, including all the available bibliographic information.

Step 6: Test and Refine Vocabulary. This is the toughest and longest phase in the development of any coordinate index. The index vocabulary, like the retrieval system itself, must be tailored to the users' needs. One of the problems in testing and refining an index vocabulary is finding the right people to do the job. Ideally the individuals should have a thorough knowledge of the subject matter field plus traning and experience in indexing. An acceptable substitute is the team approach in which professional people with knowledge of the subject are brought together. The testing and refining phase should cover at least 500 documents or a 6-month period, whichever occurs first. During this phase the temporary file should be used for actual searching, with tests made to determine the effectiveness of the vocabulary. Below are some of the things to look for and do in the testing and refining process. Figure 45 illustrates these points.

Broad or precise terms? The proper degree of indexing depth or specificity is governed by the size of the collection and user needs and can be arrived at only through a continuing analysis of these needs and system performance. In developing an index vocabulary, at the beginning one should lean toward use of broader terms in preference to the more specific terms until there is a proven need for the latter. The following are key criteria for determining how specific individual indexing terms should be:

- The terms ordinarily need be no more precise than those used in the material being indexed and by the users in their search requests. (Broad terms should ordinarily be used in areas of peripheral interest.)
- If the term receives heavy usage in indexing and heavy usage in searching and as a result more documents are retrieved than the users need or want, it probably should be replaced or supplemented by a more specific term. (It may still be necessary to retain the term in order to be able to conduct generic (general) searches.)
- If a term receives extremely light usage in indexing and searching, it probably should be dropped and included within the definition of another term, unless it is so unique or significant that it warrants retention as a separate term.

Single word or compound terms? In the early coordinate indexing systems individual terms consisted of a single word; however, it soon became apparent that there were times when two of the words should be joined. Words are joined together for one of the following reasons:

- They usually appear together in the document or form a single concept, for example, "North America," or "information retrieval."
- To provide specificity as in "metal tubing," "plastic tubing," etc.
- To prevent false retrieval caused by improper association of terms during the search process, for example, retrieval of a document about a "dog house" when the search concerned a "house dog."

While some combining of terms is necessary and beneficial, excessive or indiscriminate combining tends to defeat the basic purpose of coordinate indexing. It may result in loss of information at the time of retrieval and will increase the size of the index vocabulary.

General terms needed? Coordinate indexing, as explained earlier, is based on the principle of assigning numerous interdependent indexing terms which, when considered as a group, form a fairly complete description or, in effect, a limited abstract of the document. If the same indexing term is used for indexing documents that deal with a narrow aspect of a document and also for those that discuss the term in general, both types of documents will be retrieved if that term alone is used when conducting a search.

If searching for general documents under any particular indexing term is commonplace and re-

TESTING AND REFINING VOCABULARY TERMS

BROAD OR PRECISE TERMS?

BROAD	PRECISE
ANIMAL	Antelope Bear Cat Dog
FARMING	Cultivate Fertilize Irrigate Pollinate

SINGLE OR MULTIPLE WORDS?

SINGLE	MULTIPLE
New York	New York
Management Paperwork	Paperwork Management
Dog House	Dog House House Dog

GENERAL TERMS NEEDED?

SPEED
STOPOVER
SUPERCARGO
TRAILER
TRUCK
*TRUCKS, GENERAL
UNDERCARRIAGE
URBAN
UTILITY
WAGE
WEIGHT
WILDCAT STRIKE

Figure 45

sults in the retrieval of a large number of unwanted documents, some adjustment to the system may be necessary. One way would be to set up two indexing records for the term, one to be used when a document represents a general discussion of the terms and the other when the term is used in combination with other terms.

There are numerous other techniques for accomplishing this adjustment, including placing an asterisk beside the document number whenever it represents a general discussion of the term. However, as explained earlier, it is sometimes more practical to do a little extra screening of the output for the purpose of deleting unwanted docu-

ments than it is to try to improve system performance through additional refinements to the input process.

Step 7: Prepare the Indexing Manual. Even the simplest coordinate index system needs a manual. To make certain that the vocabulary is used as intended, it is necessary to put in writing the indexing rules, term definitions, and cross-references and to include appendixes of special reference aids needed. Indexing manuals go by many names, but all have one thing in common—they are the main control device of the coordinate index system. To the indexer, the manual is the system's "bible"; to the searcher and the user, it is an essential reference tool.

The index manual should serve as a translating tool for reconciling differences in the terms used by the authors and the users as well as to bridge the gap between the indexers and the searchers. This is accomplished by including all likely terms in the alphabetical listing of indexing terms and cross-referencing them to the equivalent terms used in the system.

It may be possible in a very small system to get by with a simple glossary, authority list, or dictionary of terms that includes definitions, where needed, and cross-references for synonyms. In the larger systems, where the indexing terms number in the hundreds or thousands, it becomes essential to know and display the relationships among the indexing terms—upward, downward, and horizontal. To answer this need, thesaurus-type indexing manuals are now in common use.

Construction of the thesaurus of indexing terms. Figure 46 shows a sample page from a thesaurus. The following is an explanation of the various headings:

Main index terms. These are the actual terms used for indexing documents. These same terms appear in the index file and constitute the index vocabulary of the system. Indexing terms consisting of two or more words should usually be listed by direct entry in their natural order; for example, RECORDS MANAGEMENT, not MANAGEMENT, RECORDS. In order to distinguish the various meanings of homographs, such qualifying expressions as TANKS (WEAPON) and TANKS (CONTAINER) may be used, in which event the qualifying expression becomes a part of the indexing.

Scope note. A short explanation used when needed to convey the meaning of an indexing term. A precise dictionary definition should not be attempted. The scope note merely indicates how the subject index term should be used and is not part of the subject index term:

> COMBUSTION CHAMBER GASES. The gases in a combustion chamber before or after ignition; for studies of gases ejected from the combustion chamber, see EXHAUST GASES.

Use reference (USE). The USE reference is intended to lead users of a thesaurus to appropriate subject index terms and should be employed to refer from a term that is not selected to one that is; for example:

1. To indicate a preferred synonym:

   SECONDARY BATTERIES *USE* STORAGE BATTERIES

2. To refer from a specific term to a more general term that has been selected to represent the specific concept:

   PLANT WAXES *USE* WAXES

   SAND BLASTING *USE* ABRASIVE BLASTING

3. To indicate a preference between spelling variations or to expand or explain abbreviations:

   INFLAMMABILITY *USE* FLAMMABILITY

   PENTAERYTHRITOL TETRANITRATE *USE* PETN

   EEG *USE* ELECTROENCEPHALOGRAMS

4. To express concepts that can be considered synonyms for purposes of indexing and retrieval:

SAMPLE PAGE FROM THESAURUS OF INDEXING TERMS

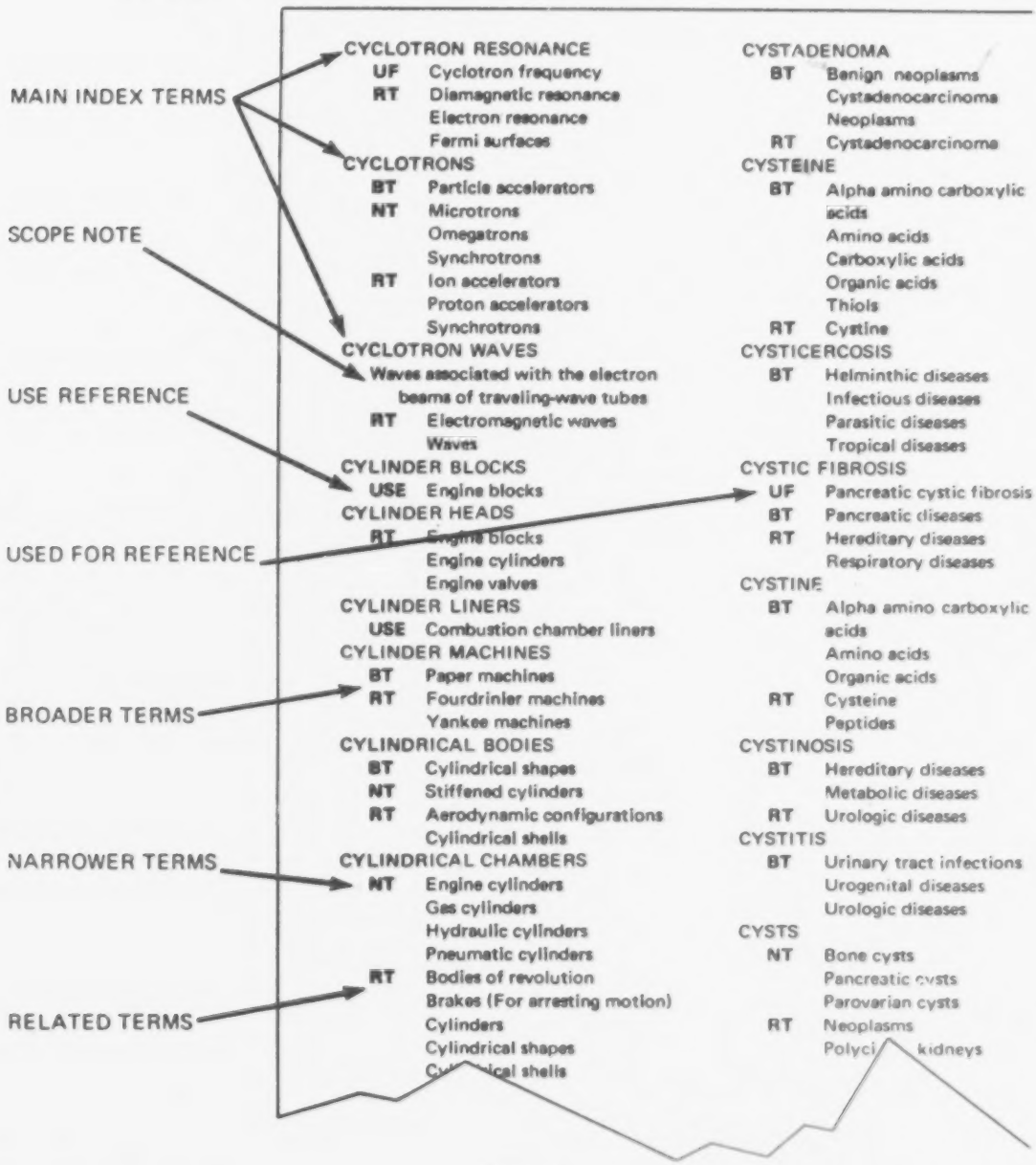


Figure 46

HEREDITY *USE* GENETICS

5. To bring together different viewpoints of a conceptual continuum:

SMOOTHNESS *USE* ROUGHNESS

6. To explain variations in word order:

TABLES (MATHEMATICS) *USE* MATHEMATICAL TABLES

7. To show how two or more index terms can be assigned to express a word not included in the index vocabulary:

HEN *USE* FEMALE AND CHICKEN

Used for reference (UF). The UF reference is the reciprocal of the USE reference. It should be used because it is essential for recordkeeping.

STORAGE BATTERIES *UF* SECONDARY BATTERIES

Broader terms (BT) and narrower terms (NT). The BROADER TERM (BT) and NARROWER TERM (NT) cross-references are employed to indicate class relationships that may exist among subject index terms. The reference is used to refer from a term symbolizing a concept class to all terms symbolizing concepts that are members of that class. The reference is used to refer from a term representing a member of a class of concepts to the term in the thesaurus representing that class. Whenever either of these cross-references is used, the reciprocal reference is also entered:

STEELS *BT* IRON ALLOYS

IRON ALLOYS *NT* STEELS

Related term (RT). The RT cross-reference is employed to refer from a subject index term to any other terms that are closely related conceptually but not hierarchically. For recordkeeping purposes, RELATED TERM references should always be entered reciprocally:

ORES *RT* MINERALS

MINERALS *RT* ORES

Hierarchical reference aids. Earlier discussions have disclosed the problems and limitations of trying to organize large bodies of complex, changing material in a hierarchical classification basis for retrieval by subjects. However, the change to a coordinate index does not eliminate the need or desirability for being able to determine hierarchical "family tree" relationships among terms. It is a natural inclination of many people to classify and organize information and items hierarchically because this is the method most familiar to them; consequently, they prefer that the reference aid be organized in this manner. Hierarchical classification schemes have their own "built-in" logic that helps the system designer, the indexer, the searcher, and the user get an overall picture of the coverage and scope of the collection and the depth of indexing.

Some manuals, therefore, also include hierarchical finding aids in which terms appearing in the straight alphabetical listing are arranged hierarchically. (See figure 47). Since these finding aids in no way change the structure of the actual vocabulary or the arrangement of the index file, several different ones can be developed, if needed, to reflect the preferences and needs of various types of user groups.

Staffing

It is futile to attempt to establish an information retrieval system without competent personnel. Otherwise, the best designed system will not be effective and a weak system may not survive long enough to give the designer an opportunity to correct the design deficiencies. A key question in planning personnel needs is: "Should subject matter specialists or professional indexers be secured?" In systems for retrieval by subjects, the need for a thorough knowledge of the subject field and the art of indexing are probably of equal importance. If a choice must be made between candidates who have only one of these skills, it is usually better to select the person who has a thorough knowledge of the subject field and then train him to be an indexer. An exception to this would be a situation where the system is used for storage and retrieval of routine general material such as newspaper clippings, in which case it should be possible, with the aid of a good indexing manual, to train any reasonably intelligent person to do the job.

Current Awareness Services

In addition to retrieving documents or data upon demand (retrospective searching), three other services that are sometimes incorporated in a coordinate indexing system are issuance of document announcement bulletins, abstracting, and selective dissemination of information. These types of current awareness services are designed to inform potential users of information about the availability and contents of recently received documents.

Announcement Bulletins. Printed periodical announcement bulletins are issued in situations where there are a large number of user groups. They list in numerical sequence descriptive infor-

EXAMPLES OF HIERARCHICAL FINDING AIDS FOR A THESAURUS

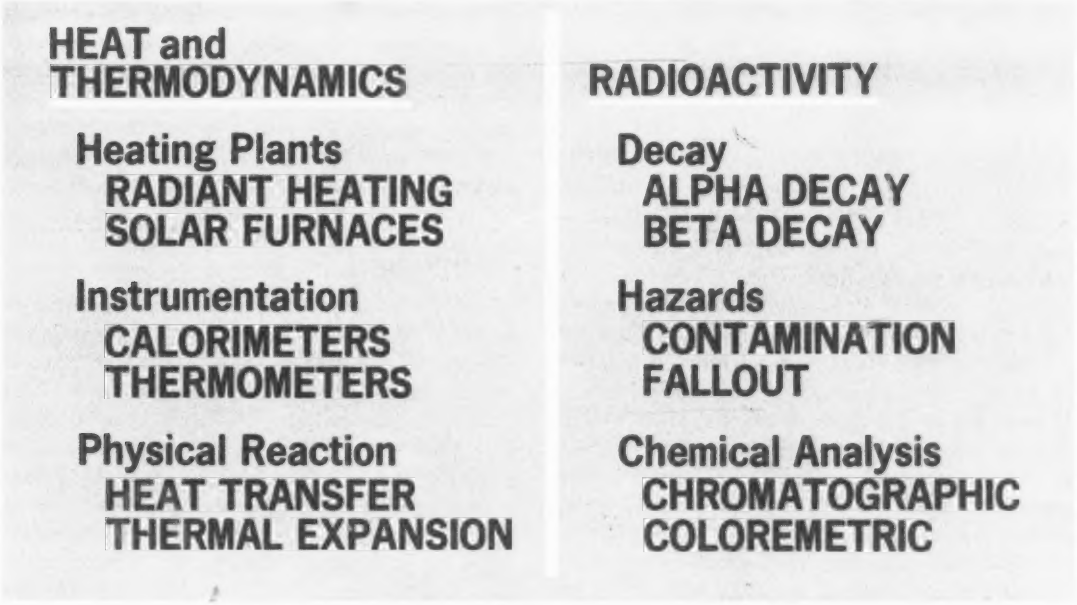


Figure 47

mation (abstracts) on newly acquisitioned documents at an information center and usually include subject or author indexes to aid in finding particular documents listed. See figure 48 for a sample of such a bulletin. Even in small information facilities, where formal published bulletins are not warranted, some method is needed to keep users informed of the availability of new accessions.

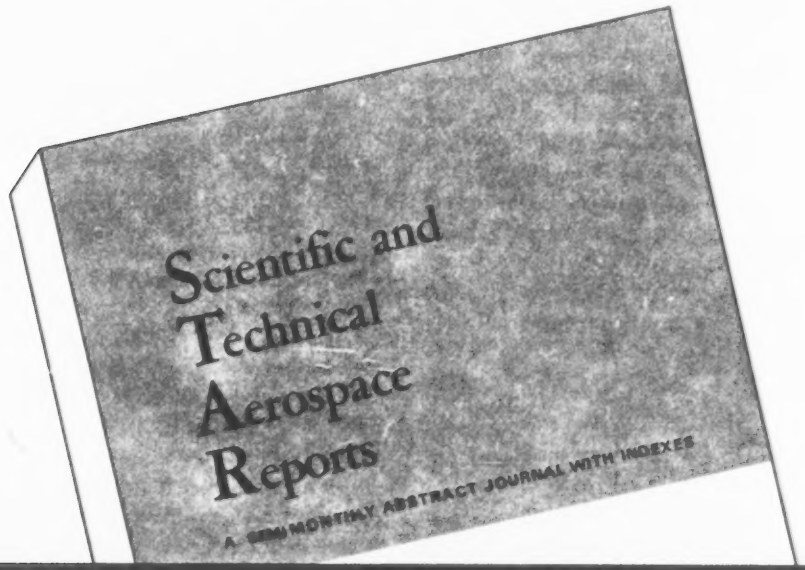
Abstracts. Because abstracting is expensive, its use should normally be restricted to situations where the documents receive widespread distribution or use. Many documents received from outside sources may already include abstracts that may be incorporated in the system at little or no expense. Most abstracts are prepared by professional indexers and editors; however, there is a growing tendency to require the authors to prepare the abstracts, a practice which in a few instances has met with failure, yet in other instances has been successful. Figure 48 also includes samples of abstracts of newly accessioned documents at the Scientific and Technical Information Facility of National Aeronautics and Space Administration. Both author- and indexer-prepared abstracts are included in this system.

Author abstracting should be given serious consideration in such fields as law, medicine, and others where case histories and decisions need to be disseminated and recorded for future study. Some professional assistance and editing may still be required, of course. Perhaps one of the greatest values of abstracts lies in their potential future use as input for automatic indexing and machine retrieval of documents.

Selective dissemination of information (SDI). As explained in chapter V, SDI involves notifying the user (or user groups) individually each time a document is received which is of the type the user has indicated might be of interest to him. To accomplish this, each user's interest profile is developed, with his help, and often maintained on computer tape. The computer compares the indexing description of each new incoming document against the user interest profiles; if they match, the computer prepares a notice that is sent to the user. The notice usually contains a description of the document and the user is given the opportunity to borrow or acquire a copy.

Effect of current awareness services. The use of announcement bulletins, selective dissemi-

SAMPLE OF A PUBLISHED ANNOUNCEMENT BULLETIN



18-13 GEOPHYSICS

N71-28798# Applied Physics Lab., Johns Hopkins Univ., Silver Spring, Md.
ALGORITHM TO COMPUTE TROPOSPHERIC REFRACTION EFFECTS ON RANGE MEASUREMENTS
S. M. Yionoulis. Jul. 1970. 13 p. refs.
(Contract N00017-62-C-0604)
(AD-721333; APL-TG-1125) Avail: NTIS CSCL 4/1

H. S. Hopfield has modeled a two-quartic tropospheric refractivity profile for correcting satellite range and range rate data which treats the dry and wet components of the tropospheric refractivity separately. The expression given for computing the contributions to range data is sensitive to rounding errors at high elevations even when evaluated in double precision. Alternate forms of the solution are presented here which eliminate the problem of rounding errors and the need for double precision computation. They allow the user to benefit from the full accuracy of Mrs. Hopfields model for all elevations. Author (GRA)

N71-28853# National Oceanic and Atmospheric Administration, Washington, D.C. Federal Coordinator for Meteorological Services and Supporting Research.
FEDERAL PLAN FOR AIR POLLUTION CONTROL METEOROLOGICAL SERVICE: TO SUPPORT FEDERAL, STATE, AND LOCAL AIR POLLUTION CONTROL AGENCIES
Jan. 1971. 27 p.
(COM-71-00200; NOAA-71012801) Avail: NTIS CSCL 13B

The needs of federal, state, and local air pollution control agencies for specialized meteorological support are investigated. Agency responsibilities, service concepts, and a 5-year program directed toward providing improved services to these users are described. Author (GRA)

N71-28866 National Lending Library for Science and Technology, Boston Spa (England).
FLUCTUATIONS OF THE TOTAL OZONE CONTENT OF THE ATMOSPHERE IN CONNECTION WITH STRATOSPHERIC WARMINGS
W. Hoebbel. 24 Jun. 1970. 9 p. refs. Transl. into ENGLISH from Met. Dienst DDR, 75 Jahre Met. Obs. Potsdam 1892 - 1967 (Potsdam), 1969 p. 108 - 111
(NLL-M-9270-(5828.4F)) Avail: Natl. Lending Library, Boston Spa, Engl. 1 NLL photocopy coupon.

Monthly means of total ozone content averaged over all stations in the Northern Hemisphere were always greater in magnitude in years in which there was a late final warming than in years with an early warming. A relationship between motion direction of sudden stratospheric temperature change regions and wind direction at the equatorial region and a high total ozone content over the Northern Hemisphere is established. G.G.

N71-28867*# Translation Consultants, Ltd., Arlington, Va.
CHANGE IN THE ELASTIC PARAMETERS AND STRENGTH OF ROCKS UNDER PRESSURE [OB IZMENENII UPRUGIKH PARAMETROV I PROCHNOSTI GORNYKH POROD POD DEYSTVIYEM DAULENIYA]
Z. I. Stakhovskaya. Washington. NASA. Jun. 1971. 8 p. refs. Transl. into ENGLISH from the publ. 'Problemy Mekhaniki Gornykh Porod, Vsesoyuznoy Nauchnoy Konferentsii, 1st' Alma-Ata, 1965 p. 394 - 398
(Contract NASw-2038)
(NASA-TT-F-13653) Avail: NTIS CSCL 08G

Models are used to demonstrate the effect porosity in rocks has on change in elastic parameters when rocks are subjected to high uniform pressures. It is shown that the change can be found by knowing the relationship between areas of contacts and units strain, and that this latter relationship can be found by using certain formulas and curves, both of which are included. Author

Figure 48

nation, abstracts, and other current awareness services are a valuable means for communicating new ideas and information and can be instrumental in reducing duplication of effort. To the designer of the information retrieval system, the incorporation of such current awareness techniques in the system is of major concern, since these techniques can substantially reduce the retrospective searching workload. The more that is done in the way of current awareness, the less searching that is required, usually. Also, unless users are kept informed and given a simple, easy method for obtaining current information, they are likely to turn to their colleagues for the needed information or to their personal files rather than to use the information facility.

Quality Control

It is necessary to achieve acceptable quality in every retrieval system, but the art is fraught with too many problems to ever be perfect. The term "quality" as used here refers to the percent of recall and precision and the absence of errors and inconsistencies.

Recall. Recall represents the percent of pertinent documents known to be in the collection that are retrieved in response to a search question. If a system has high recall, it means that only a few pertinent documents are being missed or overlooked when subject searches are made. Low recall, on the other hand, means that a substantial percentage of pertinent documents are not being retrieved.

Precision (or Relevance). Precision represents the percent of documents retrieved during a subject search that are relevant to the search question. If a system has high precision, it means that the users find that only a few irrelevant documents are being retrieved. Low precision, on the other hand, means that a large percentage of the documents retrieved are not pertinent to the search question.

Errors and Consistency. Indexing errors and lack of consistency are another major cause of indexing systems failures.

Setting Quality Standards

It is just about impossible to achieve 100 percent in both recall and precision. Improvements in recall tend to decrease precision and vice versa. However, system performance can be improved by various means. The all-important thing to remember is that management should decide what standards it wants the system to achieve; i.e., high recall and low precision, low recall and high precision, or somewhere in between. The higher standards require more costly controls, and management must weigh the value of different levels of performance in the light of the costs of achieving these levels.

Achieving Higher Recall Performance. These are the various ways that recall performance can be increased:

Harden vocabulary by careful treatment of synonyms, avoiding the use of vague terms, developing clear definitions, using standard technical terms in preference to popular jargon, and using root words to cover all variations of a term.

Use broader terms in both the vocabulary and in the assignment of terms to individual documents.

Assign more terms per document so that those topics or concepts only slightly involved are also included in the index descriptions.

More exhaustive searching by broadening the search and improving the search strategy.

Improving Precision Performance. These are some actions that can improve precision performance:

Increase vocabulary specificity by working closely with the users to develop terms that will express the needed information more precisely.

Add weights to each term assigned to the document. For example, a "1" following an index term "corrosion" might mean that the document contains information of major importance on that topic; a "2" might signify moderate importance; and a "3," minor importance. Or an asterisk could be placed in front of a term whenever it is of major importance.

Increase search specificity by having the searchers work more closely with the users in negotiation of the search in order to select more precise terms.

Reducing Errors and Increasing Consistency. These are various ways to reduce errors and increase consistency:

Training. Develop a systematic plan for training new indexers and searchers and refresher courses for experienced employees.

Prescriptive indexing. Wherever possible, prescribe in the indexing manual which term will be used in situations where there are various possibilities, instead of leaving the choice to the indexer.

Indexing and searching aids. Develop hierarchical or other "lead-in" vocabularies as an appendix to the indexing manual; also develop written rules for search strategy.

Personnel rotation. Rotate personnel between indexing and searching; also consider rotating personnel within the indexing group.

Spot checks. Use spot checks or random sampling quality control techniques. (Complete review of all indexing work would normally be too costly and usually does not completely solve the problem anyway.)

Conclusion

This chapter makes it quite clear that designing and operating a coordinate index is a formidable task. However, coordinate indexing systems offer the most powerful technique yet developed for manually organizing information and retrieving it and are essential to meet many of today's complex information needs. The only other possibility is the automatic indexing and searching system described in chapter V, which is, in effect, a form of coordinate indexing. The theory of automatic indexing is about as old as coordinate indexing; however, its development and growth have been much slower, largely due to the higher initial and input costs and the shortage of people having experience in the field.

It should also be understood that there is no such thing as a finished design for a coordinate index system. Most systems will require substantial revisions in a year or two after being established, and major revisions will occur approximately every five years. Consequently, a systems designer intimately familiar with the system should be available periodically to evaluate the performance of the system and develop plans for making the changes.

APPENDIX A

INFORMATION RETRIEVAL NONCONVENTIONAL METHODS AND EQUIPMENT GUIDE

PART A - INPUT AND STORAGE

	CHARACTERISTIC OR FEATURE	CLUE-WORD EXTRACT CARD				PERMUTED INDEX				COLUMNAR CARD				DUAL DICTIONARY			
		DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
		X				X				X				X			
PHYSICAL	MAXIMUM DOCUMENT SIZE																
	IDEAL DOCUMENT (OR DATA ELEMENT) LENGTH (1)																
	MAXIMUM DOCUMENT (OR DATA ELEMENT) LENGTH (3)																
	RESTRICTIONS ON FORM OF INFO THAT MAY BE STORED	None				Written info only				Index terms and related document numbers only				Index terms and related document numbers only			
FILE SIZE	IDEAL TOTAL NUMBER OF DOCUMENTS (OR DATA ELEMENTS) (3)	INA				500 to 2000 per listing				500 to 5000				Under 1000			
	IDEAL TOTAL VOLUME OF INFO																
	LIMITATIONS ON FURTHER EXPANSION	None except possibly space				Requires multiple sets or reissue				Ease and speed of retrieval				Ease and speed of retrieval			
INTELLECTUAL	IDEAL AMOUNT OF DESCRIPTIVE DATA PER DOCUMENT (OR DATA ELEMENT) (3)	Less than half a page				Up to 80 characters				3 to 6 index terms				3 to 6 index terms			
	IDEAL TOTAL AMOUNT OF DESCRIPTIVE DATA PER SYSTEM	INA				40,000 to 160,000 per listing				50 to 500 terms				50 to 500 terms			
	CAPABILITY - EXPANSION OF DESCRIPTIVE OR OTHER DATA	Excellent				Very limited				Excellent				Excellent			
	CAPBILITY - REORGANIZATION OF DESCRIPTIVE OR OTHER DATA	Excellent				Not necessary				Excellent				Excellent			
	CAPABILITY - CHANGING DOCUMENT ARRANGEMENT																
	SPECIAL SKILLS REQUIRED FOR INPUT PROCESSING (1)	None				Keypunch and ADP Operators				None				None			
SOURCE	CAPABILITY - ACCEPTING DATA IN MACHINE LANGUAGE	None				Excellent				Only if maintained on computer				Excellent			
	CAPABILITY - PRODUCING SYSTEM BY MEANS OF A COMPUTER	Good				Essential				Good, but requires reissue to update				Essential			
	NEED FOR STANDARDIZED FORM AND FORMAT FOR INPUT	No special requirements				Essential				No special requirements				No special requirements			
CHANGE	CAPABILITY - ADDING TO DESCRIPTIVE OR OTHER STORED DATA	Excellent				only if reissued				Excellent				Only if reissued			
	CAPABILITY - CHANGING DESCRIPTIVE OR OTHER STORED DATA	Good				Only if reissued				Good, but time consuming				Only if reissued			
	CAPABILITY - ADDING TO CONTENTS OF STORED DOCUMENTS																
	CAPABILITY - CHANGING CONTENTS OF STORED DOCUMENTS																

ABBREVIATIONS AND SYMBOLS

INA Information not available
D&D See Definitions & Descriptions

1 Refers to machine or equipment skills, only. In addition, subject matter knowledge on a par with that of users may be needed.

2 Refers to general accuracy when that class of method or equipment is used. Does not take into consideration human factors affecting accuracy or quality of the results.

3 Quantities shown for EAM Punched Card, Computer, and Computer Mass Memory refer to data elements rather than the entire document or record.

EDGE NOTCHED CARD				OPTICAL COINCIDENCE				EAM PUNCHED CARD ③				MISC. CARD SELECTORS				MICROFILM JACKET				MICROFICHE			
DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
X			-X	X			-X	X			-X		X				X				X		
												5 x 8, no limit if on microfilm				14" wide, any length or 40 x 60"				Normally 8½ x 11"			
												Same as that for record medium				9 to 60 pages				9 to 60 pages			
												Same as that for record medium				Impractical for over 120 pages				Impractical for over 120 pages			
None				Index terms and related document numbers only				Alphanumeric characters and symbols				Same as that for record medium				None				None			
1000 to 5000				1000 to 10,000				DR 1000 to 20,000 DFR 1000 and up				1000 cards and up				5000 and up				10,000 and up			
								DR-N/A, DFR up to 1,500,000				1000 pages or sheets and up				50,000 pages and up				100,000 pages and up			
Ease and speed of retrieval				Multiple setslowers search speed				Lower retrieval speed				None				None				None			
3-6 items of coded data, or half page of other info				3 to 10 index terms				Under 80 characters				Varies, usually under 15 digits				Under 45, but OK up to 180				Under 45, but OK up to 180			
No ideal amount				50-1000 index terms				80,000 up				No ideal amount				45-180 x the no. of documents				45-180 x the no. of documents			
Limited				Satisfactory				Excellent, if the space permits				None, as far as the selector is concerned				OK up to 180 Char.				OK up to 180 Char.			
Limited				Excellent				Excellent				None, as far as the selector is concern				Possible, but time consuming				Usually requires refilming			
												Not necessary due to random filling				Satisfactory				Satisfactory			
None				None				Keypunch and ADP operators				None				Microfilming				Microfilming			
None				Some systems are excellent				Excellent				Usually none				None				None			
None				Some systems are excellent				Excellent				Same as that for record medium				Excellent				Presently limited			
Essential				Essential				Essential				Essential				No special requirements				Essential			
Only if the space permits				Excellent				Excellent, if the space permits				Same as that for record medium				Good				Limited			
Very limited				Very limited				Excellent, but time consuming				Same as that for record medium				Limited				Requires refilming or index, usually			
												Same as that for record medium				Excellent				Requires refilming or index, usually			
												Same as that for record medium				Limited				Requires filming or index, usually			

PART A - INPUT AND STORAGE *(Continued)*

	CHARACTERISTIC OR FEATURE	MICROFILM STRIP				MICROFILM ROLL MECHANIZED				MICROFILM ROLL PHOTO-OPTICAL				MICROFILM CHIP AUTOMATED			
		DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
			-X		-X		X		-X	X	X	X	-X		X	-X	
PHYSICAL	MAXIMUM DOCUMENT SIZE	14" wide, any length or 40 x60"				14" wide any length				14" wide any length				8½ x 14" or 40 x60"			
PHYSICAL	IDEAL DOCUMENT (OR DATA ELEMENT) LENGTH ③	DS - up to 40 pages DFR-25 to 50 lines				DS - one page DFR-50 lines				DR-any; DS/URS 1-8 pages; DFR-25-50 lines				DS - 1 to 2 pages URS-1 to 8 pages			
PHYSICAL	MAXIMUM DOCUMENT (OR DATA ELEMENT) LENGTH ③	DS-none DFR-1000 lines				Not necessary if over 100 pages or 5000 lines				DR-none; DS/URS/ DFR-not nec. if over 25 pages				Should not exceed above			
PHYSICAL	RESTRICTIONS ON FORM OF INFO THAT MAY BE STORED	None				None				None				None			
FILE SIZE	IDEAL TOTAL NUMBER OF DOCUMENTS (OR DATA ELEMENTS) ③	1000 and up				10,000 and up				25,000 and up				DS-50,000 and up URS-100,000 and up			
FILE SIZE	IDEAL TOTAL VOLUME OF INFO	DS - up to 40,000 DFR-25,000 lines and up				DS - 10,000 and up DFR - 250,000 lines and up				DS/URS-25,000 and up DFR-million lines and up				DS-50,000 pages up URS-500,000 up			
FILE SIZE	LIMITATIONS ON FURTHER EXPANSION	None				None				URS-speed and cost others- none				None			
INTELLECTUAL	IDEAL AMOUNT OF DESCRIPTIVE DATA PER DOCUMENT (OR DATA ELEMENT) ③	Under ten characters				Under ten characters				Varies-9 digits or 56 characters				DS-up to 18 digits URS-up to 100 Char.			
INTELLECTUAL	IDEAL TOTAL AMOUNT OF DESCRIPTIVE DATA PER SYSTEM	10,000 char. and up				100,000 characters and up				200,000 digits or characters and up				DS-900,000 digits up URS-10 M and up			
INTELLECTUAL	CAPABILITY - EXPANSION OF DESCRIPTIVE OR OTHER DATA	Limited and may require refilming				Limited and may require refilming				Requires refilming				DS-requires refilming URS-OK up to 100			
INTELLECTUAL	CAPABILITY - REORGANIZATION OF DESCRIPTIVE OR OTHER DATA	Requires refilming				Requires refilming				Requires refilming				DS-requires refilming URS-excellent			
INTELLECTUAL	CAPABILITY - CHANGING DOCUMENT ARRANGEMENT	Requires refilming				Requires refilming				Requires refilming				Good			
INTELLECTUAL	SPECIAL SKILLS REQUIRED FOR INPUT PROCESSING ①	Microfilming				Microfilming				Microfilming and ADP type skills				Microfilming and ADP type skills			
SOURCE	CAPABILITY - ACCEPTING DATA IN MACHINE LANGUAGE	None				None				Limited				DS-none URS-excellent			
SOURCE	CAPABILITY - PRODUCING SYSTEM BY MEANS OF A COMPUTER	Excellent				Excellent				16 MM, excellent				Partial			
SOURCE	NEED FOR STANDARDIZED FORM AND FORMAT FOR INPUT	DS-none DFR-essential				Essential				Essential				Essential			
CHANGE	CAPABILITY - ADDING TO DESCRIPTIVE OR OTHER STORED DATA	Requires slicing, refilming, or index				Requires splicing, refilming, or index				Requires splicing, refilming, or index				DS-requires refilming URS-excellent			
CHANGE	CAPABILITY - CHANGING DESCRIPTIVE OR OTHER STORED DATA	See above				See above				See above				DS-requires refilming URS-excellent			
CHANGE	CAPABILITY - ADDING TO CONTENTS OF STORED DOCUMENTS	See above				See above				See above				Satisfactory by adding new chips			
CHANGE	CAPABILITY - CHANGING CONTENTS OF STORED DOCUMENTS	See above				See above				See above				Requires refilming			

MICROFILM EAM PUNCH CARD				MICROFILM NOTCHED CARD				MICROFILM SUPERMINIATURE				VIDEO TAPE				COMPUTER-ALL ③				COMPUTER MASS MEMORY ①			
DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
	X	X			X	X			X				X		-X	X			X	X			X
14" wide, any length or 40 ×60"				14" wide, any length, or 40 × 60				INA				Up to 8½ × 14"											
1 to 8 pages				1 to 8 pages				INA				DS - one page DFR-25-50 lines				Up to 100 characters				Up to 100 characters			
Should not exceed above				Should not exceed above				INA				Cost and efficiency loss				None				None			
None				None				None				None				Numeric, alpha and special characters				Numeric, alpha and special characters			
DS-10,000 and up URS-1000-20,000				1000 - 5000				Undetermined				200,000 and up				20,000 and up no fixed ideal				Over 1,000,000 on line			
DS-10,000 pages up URS-1K - 160,000 page				1000-40,000 pages				Undetermined				DS-200,000 pages up DFR-5million lines				200,000 up no fixed ideal				Over 100,000,000 on line			
DS - none URS- search time				Speed				Undetermined				Speed and cost				On line physical Off line-time				On line-physical Off line-time			
Up to 58 characters				3 - 6 coded items and half page of other				Undetermined				Up to 18 numeric or 12 alpha characters				Up to 100% of the ideal document length				Up to 100% of the ideal document length			
DS-58,000 char. up URS-50,000 to 1M				No ideal amount				Undetermined				2 million characters up				See above				See above			
Excellent, if space permits				Excellent, if space permits				Undetermined				Excellent, but may be wasteful of space				Limited only by the search time and cost				Limited only by the search time and cost			
Excellent				None				Undetermined				Excellent				Excellent				Usually requires re-recording			
Excellent				Excellent				Undetermined				Good				Satisfactory, but may not be necessary				See above			
Microfilming and ADP type skills				Microfilming				Microfilming				Electronics and ADP type skills				ADP operators				ADP operators			
Good				None				None				INA				Excellent				Excellent			
Partial				None				Partial				Good				Excellent				Excellent			
Essential				Essential				Essential				Essential				Essential				Essential			
Excellent, if space permits				Only if space permits				limited to some systems only				Only if space permits				Limited only by the search time and cost				Requires a special procedure			
Excellent, but time consuming				Very limited				See above				Excellent				Excellent				Usually requires re-recording			
Satisfactory by adding new cards				Satisfactory by adding new cards				See above				Excellent by adding new pages				Limited only by the search time and cost				Requires a special procedure			
Requires refilming				Requires refilming				limited to some systems only				Good, by erasing and re-recording				Excellent				Usually requires re-recording			

PART B - RETRIEVAL AND PRESENTATION

	CHARACTERISTIC OR FEATURE	CLUE-WORD EXTRACT CARD				PERMUTED INDEX				COLUMNAR CARD				DUAL DICTIONARY			
		DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
		X				X				X				X			
ACTIVITY	IDEAL AVERAGE NO. OF RETRIEVAL ACTIONS PER DAY	INA				Up to 5 per copy				Up to 20				Up to 5 per copy			
	CAPABILITY - SIMULTANEOUS USE BY TWO OR MORE PEOPLE	Excellent				Excellent				Only if more than one set is kept				Excellent			
	CAPABILITY - LOW COST DECENTRALIZATION TO USER LOCATIONS	Good				Excellent				Good				Excellent			
	CAPABILITY - DIRECT QUERYING FROM REMOTE SITES	None				Not necessary				None				Not necessary			
	CAPABILITY - HANDLING LARGE PEAK LOADS	Excellent				Possible, but not always practical				Limited				Possible, but not always practical			
	PORTABILITY	Poor				Excellent				Good				Excellent			
INTELLECTUAL	CAPABILITY - CORRELATING AND MANIPULATING STORED DATA	See D & D				Limited				Good				Good			
	NO. OF DESCRIPTIVE TERMS THAT CAN BE SEARCHED AT SAME TIME	1				1				2				2			
	HUMAN INTERVENTION INVOLVED IN SEARCH OR LOOK-UP PROCESS	100%				100 %				100%				100%			
	NO. OF STEPS REQUIRE PER RETRIEVAL ACTION	Varies, but can be numerous				Continuous scanning				12 for each term searched				11 for each term searched			
	SPECIAL SKILLS REQUIRED IN USING THE SYSTEM (1)	None				None				None				None			
PHYSICAL	OUTPUT - TYPE OF INFO OR DATA FURNISHED BY THE SYSTEM	Descriptive data and extract				names, titles, key terms, etc.				document numbers				document numbers			
	PRESENTATION OR DISPLAY METHOD	Direct viewing				Direct viewing				Direct viewing				Direct viewing			
	CAPABILITY -RETRIEVAL OF PORTIONS, ONLY, OF DOCUMENTS																
SERVICE	RESPONSE TIME - FROM INITIATION OF QUERY UNTIL USER VIEWS RESULTS	5 minutes and up				5 minutes and up				5 minutes and up				5 minutes and up			
	RESPONSE TIME - OBTAINING PAPER COPY OF SINGLE PAGE OR SHEET																
	RESPONSE TIME - OBTAINING MICROFILM COPY																
	CONVERTABILITY TO OTHER METHODS AND EQUIPMENT	Yes, if data is kept machinable				Yes, if provided for in design				only if data is kept machinable				Yes, if data is kept machinable			
	CAPABILITY - PRODUCTION OF DUPLICATE OR ALTERNATE SYSTEM	Good				Excellent				Excellent				Excellent			
	CAPABILITY - USER SELF SERVICE	Excellent				Excellent				Excellent				Excellent			
	CAPABILITY - DIRECT BROWSING	Excellent				Excellent				Excellent				Excellent			
	CAPABILITY - USE FOR CURRENT AWARENESS	Limited				Excellent				None				None			
	PHYSICAL EASE IN USING THE SYSTEM	Tiresome if used extensively				Tiresome, if large and heavily used				Tiresome, if used extensively				Tiresome, if used extensively			
QUALITY	ACCURACY OF RESULTS (2)	INA				Satisfactory				Error prone				Satisfactory			
	EQUIPMENT RELIABILITY																
	SECURITY AND DURABILITY OF RECORDING MEDIA	Satisfactory				Satisfactory				Satisfactory				Satisfactory			
	EQUIPMENT DURABILITY																

EDGE NOTCHED CARD				OPTICAL COINCIDENCE				EAM PUNCHED CARD ③				MISC. CARD SELECTORS				MICROFILM JACKET				MICROFICHE			
DR	DS	URS	DFR	DF	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
X			-X	X			-X	X			-X		(				X				X		
Up to 10				Up to 25				Up to 15				Up to 200				100 up				100 up, total of all locations			
None				Only if more than one set is kept				None				None				Good				Good			
Good				Good				Very limited				Very limited				Good				Excellent			
None				None				None				None				None				None			
None				Limited				Limited				Limited				Good				Good			
Limited				Limited				None				None				Excellent				Excellent			
Good				Very Good				Very Good															
1 usually				As many as desired, usually				1 to 4 normally															
Extensive				Extensive				Extensive				Nominal				100%				100%			
Up to 9 for each term searched				2 for each term searched				Varies, but usually extensive				2 if the document no. is known				4 or more				4 or more			
Knowledge of needle sorting				None				ADP operations				None				None				None			
Document nos., misc. written, graphic				Document number address				Any data recorded on cards				Same as that for record medium				Page images				Page images			
Direct viewing				Visual display or printout, see D & D				EAM cards or printout listing				Same as that for record medium				Viewer screen, paper or film copy				Viewer screen, paper or film copy			
												None				Full pages usually				Full pages usually			
20 minutes up				2 minutes				20 minutes up				Less than 10 sec. if document no. is known				Less than 1 minute				Less than 1 minute			
												5-60 sec, may need an extra unit				Less than 1 minute				Less than 1 minute			
												Up to 30 sec., may need an extra unit				Less than 1 minute				Less than 1 minute			
No				Yes in some systems				Yes				To manual methods only				To microfiche or aperture card only				None			
Very limited				Excellent				Good				Same as that for record medium				Good, but time consuming				Excellent			
Limited				Excellent				Intermediary is usually required				Good, but not always practical				Satisfactory				Excellent			
Good				Excellent				None				Good, but not always practical				Satisfactory				Good			
None				None				Limited															
Tiresome, if heavily used				Very easy				Tiresome if used extensively				Excellent				Good, but subject to user resistance				Good, but subject to user resistance			
Satisfactory				Excellent				Excellent				Satisfactory				Satisfactory				Satisfactory			
Satisfactory				Satisfactory				Excellent				Satisfactory				Satisfactory				Satisfactory			
Satisfactory, usually				Very Good				Subject to wear				Same as that for record medium				Good				Very Good			
Satisfactory				Satisfactory				Excellent				INA				Satisfactory				Satisfactory			

PART B - RETRIEVAL AND PRESENTATION

	CHARACTERISTIC OR FEATURE	MICROFILM STRIP				MICROFILM ROLL MECHANIZED				MICROFILM ROLL PHOTO-OPTICAL				MICROFILM CHIP AUTOMATED			
		DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
			-X		-X		X		-X	X	X	X	-X		X	-X	
ACTIVITY	IDEAL AVERAGE NO. OF RETRIEVAL ACTIONS PER DAY	DS-under 100 DFR- 500 up				100 up				Varies - often inverse to info volume				DS-500 to 700 URS-1000 up			
	CAPABILITY - SIMULTANEOUS USE BY TWO OR MORE PEOPLE	DS- Good DFR- None				None				None				Limited			
	CAPABILITY - LOW COST DECENTRALIZATION TO USER LOCATIONS	Limited				Good				Very limited				None			
	CAPABILITY - DIRECT QUERYING FROM REMOTE SITES	None				None				None				Depends on the equipment used			
	CAPABILITY - HANDLING LARGE PEAK LOADS	Limited				Limited				Limited				OK, if batching is permissable			
	PORTABILITY	Limited				Limited				None				None			
INTELLECTUAL	CAPABILITY - CORRELATING AND MANIPULATING STORED DATA									Good, unless multi rolls are hinderance				DS- None; URS- same as the host computer			
	NO. OF DESCRIPTIVE TERMS THAT CAN BE SEARCHED AT SAME TIME									6 to 15				Same as above			
	HUMAN INTERVENTION INVOLVED IN SEARCH OR LOOK-UP PROCESS	100%				100%				Nominal				Nominal			
	NO. OF STEPS REQUIRED PER RETRIEVAL ACTION	4 or more				4 or more				3 or more				4 or more			
	SPECIAL SKILLS REQUIRED IN USING THE SYSTEM (1)	None				None				Machine searching				Microfilming and machine operations			
PHYSICAL	OUTPUT - TYPE OF INFO OR DATA FURNISHED BY THE SYSTEM	Page images				Page images				Page images				Page images			
	PRESENTATION OR DISPLAY METHOD	DS- paper copy DFR- view screen				View screen or a paper copy				View screen or a paper copy				View screen and/or paper or film copy			
	CAPABILITY - RETRIEVAL OF PORTIONS, ONLY, OF DOCUMENTS	DS- whole document DFR- full page				Full pages, usually				Full pages only				Full pages only			
SERVICE	RESPONSE TIME - FROM INITIATION OF QUERY UNTIL USER VIEWS RESULTS	DS- 30 min. up DFR- less than 30 sec.				half to 2 minutes				30 sec. plus 15 sec. per 1000 pages				DS- under 30 sec. URS- few seconds			
	RESPONSE TIME - OBTAINING PAPER COPY OF SINGLE PAGE OR SHEET					Less than 30 seconds				Less than 30 seconds				Under 30 seconds			
	RESPONSE TIME - OBTAINING MICROFILM COPY									Any copy feature, not for use of the user				Under 10 seconds, where available			
	CONVERTABILITY TO OTHER METHODS AND EQUIPMENT	None				Microfilm jackets and strips only				Usually jackets and film strips				Aperature cards, in some systems			
	CAPABILITY - PRODUCTION OF DUPLICATE OR ALTERNATE SYSTEM	Good				Good				Good				Good			
	CAPABILITY - USER SELF SERVICE	DS- very limited DFR- limited				Limited				Very limited				Very limited			
	CAPABILITY - DIRECT BROWSING	Good				Excellent				Limited				None			
	CAPABILITY - USE FOR CURRENT AWARENESS	Limited				None				None				None			
	PHYSICAL EASE IN USING THE SYSTEM	DS- awkward DFR- satisfactory				Good, but subject to user resistance				Good, but varies with diferent equipment				Nominal			
QUALITY	ACCURACY OF RESULTS (2)	Satisfactory				Satisfactory				Excellent				Excellent			
	EQUIPMENT RELIABILITY	Good				Satisfactory				Satisfactory				Satisfactory			
	SECURITY AND DURABILITY OF RECORDING MEDIA	Very good				Satisfactory				Satisfactory				Satisfactory			
	EQUIPMENT DURABILITY	Satisfactory				Satisfactory				Satisfactory				Satisfactory			

MICROFILM EAM PUNCH CARD				MICROFILM NOTCHED CARD				MICROFILM SUPERMINIATURE				VIDEO TAPE				COMPUTER-ALL ③				COMPUTER MASS MEMORY ③			
DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
	X	-X			X	X			X				X		-X	X			X	X			X
DS- 100 up URS- up to 15				DS- Up to 100 URS- up to 10				Undetermined				1000 up				Varies with data volume, service, etc.				Varies with data volume, service, etc.			
DS- excellent URS- None				DS- excellent URS- none				Very limited				None				Possible, but not always practical				Depends on the host computer			
DS- excellent URS- very limited				Excellent				Good				Limited				Very limited				Very limited			
None				None				INA				Good				Possible, but not always practical				Depends on the host computer			
DS- good URS- limited				DS- good URS- none				Limited				Limited				Often requires batching				Depends on the host computer			
DS- good URS- none				Limited				Good				None				None				None			
Very good				Good								None				Excellent				Depends on the host computer			
1 to 4 usually				1 usually								None				No limit, usually				Depends on the host computer			
Extensive				Extensive				Varies widely				Nominal				Nominal, except for remote query				Depends on the host computer			
DS- 4 or more URS- numerous				DS- 4 or more URS- 6-9 per term				3 or more				3 or more				Varies widely				Varies widely			
Microfilming and ADP operations				Knowledge of needle sorting				Microminiature tech.				Machine Operations				ADP programming				ADP programming			
Page image				Page image				Page image				Page image				Computer stored or generated data				Computer stored or generated data			
View screen, paper or film copy				View screen, paper or film copy				Usually view screen only				View screen and paper copy				Punch card, printout, and video				Punch card, printout, and video			
Full page only				Full page only				Full page only				Full page only				Excellent				Depends on the host computer			
DS- less than 1 min. URS- 20 min. up				DS- less than 1 min. URS- 20 min. up				Less than 1 min.				Varies, average is 1 minute				Varies with each system				Varies with each system			
Under 30 sec.				Under 30 sec.								Varies with print equipment				Varies with each system				Varies with each system			
Under 1 min.				Under 1 min.																			
Yes				DS- Yes URS- No				INA				None				Possible, but may be costly				Possible, but may be costly			
Excellent				Very limited				Excellent				Good				Excellent				Excellent			
DS- Excellent URS- Very limited				Limited				Excellent				Excellent				Limited, at present				Depends on the host computer			
DS- Good URS- None				Good				Good				Excellent				Limited, at present				Limited, at present			
Limited				None								None				Excellent							
DS- may be resisted URS- can be tiring				DS- may be resisted URS- can be tedious				Good, but subject to user resistance				Good, but subject to user resistance				Very good				Very good			
Excellent				Good				Satisfactory				INA				Excellent				INA			
Excellent				Good				INA				INA				Usually very good				INA			
Unit record machines may damage film				Satisfactory				Satisfactory				Image quality may drop off				Requires special care and supervision				Requires special care and supervision			
Excellent				Excellent				INA				INA				Satisfactory				INA			

PART C - RESOURCES *(to be completed by the person making the study)*

	FACTOR	CLUE-WORD EXTRACT CARD				PERMUTED INDEX				COLUMNAR CARD				DUAL DICTIONARY			
		DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
		X				X				X				X			
INITIAL COSTS	PERSONNEL TO DESIGN AND ASSIST IN INSTALLATION OF THE SYSTEM																
	PERSONNEL TO SUPERVISE SYSTEM AND PROVIDE REFERENCE SERVICE																
	PERSONNEL TO ENTER INFORMATION INTO THE SYSTEM																
	INFORMATION ACQUISITION COSTS																
	EQUIPMENT COSTS AND SERVICE CHARGES																
	SUPPLIES, SPACE, AND MISC. COSTS																
CURRENT CAPABILITY *(availability of)*	PERSONNEL TO DESIGN AND ASSIST IN INSTALLATION OF THE SYSTEM																
	PERSONNEL TO SUPERVISE SYSTEM AND PROVIDE REFERENCE SERVICE																
	PERSONNEL TO ENTER INFORMATION INTO THE SYSTEM																
	INFORMATION TO BE ENTERED INTO THE SYSTEM																
	EXTERNAL INFORMATION SERVICES																
	EXISTING MECHANIZED EQUIPMENT OR ADP SERVICES																
	EXISTING INFORMATION AIDS OR TOOLS																
	EXISTING COMMUNICATION AND TRANSPORTATION SERVICES																
	SUITABLE SPACE																
	ELECTRIC POWER																

EDGE NOTCHED CARD				OPTICAL COINCIDENCE				EAM PUNCHED CARD				MISC. CARD SELECTORS				MICROFILM JACKET				MICROFICHE			
DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
X			-X	X			-X	X			-X		X				X				X		

PART C - RESOURCES (Continued)

	FACTOR	MICROFILM STRIP				MICROFILM ROLL MECHANIZED				MICROFILM ROLL PHOTO-OPTICAL				MICROFILM CHIP AUTOMATED			
		DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
			-X		-X		X		-X	X	X	X	-X		X	-X	
INITIAL COSTS	PERSONNEL TO DESIGN AND ASSIST IN INSTALLATION OF THE SYSTEM																
	PERSONNEL TO SUPERVISE SYSTEM AND PROVIDE REFERENCE SERVICE																
	PERSONNEL TO ENTER INFORMATION INTO THE SYSTEM																
	INFORMATION ACQUISITION COSTS																
	EQUIPMENT COSTS AND SERVICE CHARGES																
	SUPPLIES, SPACE, AND MISC. COSTS																
CURRENT CAPABILITY (availability)	PERSONNEL TO DESIGN AND ASSIST IN INSTALLATION OF THE SYSTEM																
	PERSONNEL TO SUPERVISE SYSTEM AND PROVIDE REFERENCE SERVICE																
	PERSONNEL TO ENTER INFORMATION INTO THE SYSTEM																
	INFORMATION TO BE ENTERED INTO THE SYSTEM																
	EXTERNAL INFORMATION SERVICES																
	EXISTING MECHANIZED EQUIPMENT OR ADP SERVICES																
	EXISTING INFORMATION AIDS OR TOOLS																
	EXISTING COMMUNICATION AND TRANSPORTATION SERVICES																
	SUITABLE SPACE																
	ELECTRIC POWER																

MICROFILM EAM PUNCH CARD				MICROFILM NOTCHED CARD				MICROFILM SUPERMINIATURE				VIDEO TAPE				COMPUTER-ALL				COMPUTER MASS MEMORY			
DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
	X	-X			X	X			X				X		-X	X			X	X			X

APPENDIX B

INFORMATION RETRIEVAL EQUIPMENT AND SUPPLIES SOURCES

The following list is based on the best available information at the time this handbook was prepared. Any manufacturers not included will, upon notification to NARS, be included in the next revision.

MICROFILM ROLL MECHANIZED (OR MOTORIZED)

Alpha-Vector Inc.
501 Fifth Ave.
New York, N.Y. 10017

Atlantic Microfilm Corp.
700 South Main St.
Spring Valley, N.Y. 10977

Bell and Howell Co.
Business Equipment Group
6800 McCormick Rd.
Chicago, Ill. 60645

Computer Micro Viewing, Inc.
15 Tenth Ave. South
Hopkins, Minn. 55343

DASA Corp.
15 Stevens St.
Andover, Mass. 01810

Eugene Dietzgen Co., Inc.
2425 N. Sheffield Ave.
Chicago, Ill. 60614

The Ednalite Corp.
200 North Water St.
Peekskill, N.Y. 10566

Information Design Inc.
3247 Middlefield Rd.
Menlo Park, Calif. 94025

Information Handling Services, Inc.
Denver Technological Center
Englewood, Colo. 80110

Keuffel and Esser Company
30 Whippany Rd.
Morristown, N.J. 07960

Eastman Kodak Co.
Business Systems Market Div.
343 State St.
Peekskill, N.Y. 10566

Micro 8 Co.
P. O. Box 1087
La Crosse, Wis. 54601

Microfilm Products, Inc.
40 West 15th St.
New York, N.Y. 10011

Microsystems Inc.
1717 Barnum Ave.
Bridgeport, Conn. 06610

3M Co.
Microfilm Products Div.
3M Center—220-10
St. Paul, Minn. 55101

Morgan Information Systems, Inc.
193 Constitution Dr.
Menlo Park, Calif. 94025

Reproduction Systems
One California St.
San Francisco, Calif. 94111

Stromberg DatagraphiX, Inc.
P. O. Box 2449
San Diego, Calif. 92112

Washington Scientific Industries, Inc.
13111 Wayzata Blvd.
Minnetonka, Minn. 55343

and others

ROLL MICROFILM WITH PHOTO-OPTICAL BINARY CODE

AIL Information Systems (formerly FMA)
5730 Arbor Vitae St.
Los Angeles, Calif. 90045

Alpha-Vector Inc.
501 Fifth Ave.
New York, N.Y. 10017

Eastman Kodak Co.
Business Systems Market Div.
343 State St.
Peekskill, N.Y. 10566

Morgan Information Systems, Inc.
193 Constitution Dr.
Menlo Park, Calif. 94025

Stromberg DatagraphiX, Inc.
P. O. Box 2449
San Diego, Calif. 92112

and others

MICROFILM STRIP

Eastman Kodak Co.
Business Systems Market Div.
343 State St.
Peekskill, N.Y. 10566

Other microfilm strip systems are sometimes prepared in-house.

MICROFILM CHIP, AUTOMATED

Houston Fearless Corp.
11801 West Olympic Blvd.
Los Angeles, Calif. 90064

Itek Corp.
1001 Jefferson Rd.
Rochester, N.Y. 14603

Photo Devices Inc.
33 Litchfield St.
Rochester, N.Y. 14608

and others

MICROFICHE AND MICROFILM JACKETS

Arcata Microfilm
700 South Main St.
Spring Valley. N.Y. 10977

Atlantic Microfilm Corp.
700 South Main St.
Spring Valley, N.Y. 10977

Walter M. Ballard Co.
7705 Georgia Ave., NW
Washington, D.C. 20012

Bell and Howell Co.
Business Equipment Group
6800 McCormick Rd
Chicago, Ill. 60645

Business Efficiency Aids, Inc.
8114 Lawndale Ave.
Skokie, Ill. 60076

Camera Optic Mfg. Co.
23-53 Sternway St.
Long Island, N.Y. 11105

DASA Corp.
15 Stevens St.
Andover, Mass. 01810

Data Reproduction Systems
300 East Beach Ave.
Inglewood, Calif. 90302

Eugene Dietzgen Co. Inc.
2425 N. Sheffield Ave.
Chicago, Ill. 60614

DuKane Corp.
St. Charles, Ill. 60174

GAF-Reprographic Products
140 West 51st St.
New York, N.Y. 10020

Houston Fearless Corp.
11801 West Olympic Blvd.
Los Angeles, Calif. 90064

Keuffel and Esser Co.
30 Whippany Rd.
Morristown, N.J. 07960

Eastman Kodak Co.
Business Systems Market Div.
343 State St.
Peekskill, N.Y. 10566

Micro Design, Inc.
2355 Johnson St.
Hartford, Wis. 53027

Microdisplay Systems, Inc.
Subsidiary of Communication Systems
Two Penn Plaza
New York, N.Y. 10001

Micrographic Technology Corp.
1732 Kaiser Ave.
Santa Ana, Calif. 92705

Micro Image Corp.
10469 Roselle St.
San Diego, Calif. 92121

3M Co.
Microfilm Products Div.
3M Center—220-10
St. Paul, Minn. 55101

Minolta Corp.
200 Park Ave. South
New York, N.Y. 10003

Mosler Co.
1561 Grand Blvd.
Hamilton, Ohio 45012

National Capital Systems, Inc.
P.O. Box 3762
Washington, D.C. 20007

NB Jackets
31-31 31st St.
Long Island City
New York, N.Y.

NCR-Industrial Products Div.
3100 Valleywood Dr.
Dayton, Ohio 45429

Photo Devices Inc.
33 Litchfield St.
Rochester, N.Y. 14608

Frederick Post Co.
P. O. Box 803
Chicago, Ill. 60690

Randomatic Data Systems Inc.
344 West State St.
Trenton, N.J.

Readex Microprint Corp.
5 Union Square
New York, N.Y. 10003

Realist, Inc.
N93 W16288 Mogal Dr.
Menomonee Falls, Wis. 53051

Remington Rand Office Systems
P. O. Box 171
Marietta, Ohio 45750

Sanders Associates, Inc.
1 Fairchild Ave.
Plainview, N.Y. 11802

Stromberg DatagraphiX, Inc.
P. O. Box 2449
San Diego, Calif. 92112

The Taylor-Merchant Corp.
25 West 45th St.
New York, N.Y. 10036

Technifax Corp.
195 Appleton St.
Holyoke, Mass. 01042

University Microfilms, Inc.
300 North Zeeb Rd.
Ann Arbor, Mich. 48103

Visu-Flex Co.
633 South Carondelet
Los Angeles, Calif. 90057

Washington Scientific Industries, Inc.
13111 Wayzata Blvd.
Minnetonka, Minn. 55343

and others

APERTURE CARDS (MICROFILM—EAM PUNCHED CARD OR MICROFILM—EDGE-NOTCHED CARD)

Advanced Technology Corp.
P. O. Box 246
Chambersburg, Pa. 17201

DASA Corp.
15 Stevens St.
Andover, Mass. 01810

Dakota Microfilm Service
9655 W. Colfax Ave.
Denver, Colo. 80215

Eugene Dietzgen Co., Inc.
2425 N. Sheffield Ave.
Chicago, Ill. 60614

B. K. Elliott Co.
P. O. Box 3240
Pittsburgh, Pa. 15230

Itek Business Products
1001 Jefferson Rd.
Rochester, N.Y. 14603

Keuffel and Esser Co.
30 Whippany Rd.
Morristown, N.J. 07960

Kleer-Vu Industries, Inc.
878 Sussex Blvd.
Broomall, Pa. 19008

Eastman Kodak Co.
Business Systems Markets Div.
343 State St.
Rochester, N.Y. 14650

Litton Automated Business Systems
1700 Wisconsin Ave., NW
Washington, D.C. 20007

Microseal
2222 West Main St.
Evanston, Ill. 60204

3M Co.
Microfilm Products Div.
3M Center—220-10
St. Paul, Minn. 55101

Mosler Co.
1561 Grand Blvd.
Hamilton, Ohio 45012

Remington Rand Office Systems
P. O. Box 171
Marietta, Ohio 45750

The Taylor-Merchant Corp.
25 West 45th St.
New York, N.Y. 10036

Washington Scientific Industries, Inc.
13111 Wayzata Blvd.
Minnetonka, Minn. 55343

and others

SUPERMINIATURE MICROFORMS

Microform Data Systems, Inc.
2700 Sand Hill Rd.
Menlo Park, Calif. 94025

NCR-Industrial Products Div.
3100 Valleywood Dr.
Dayton, Ohio 45429

and others

VIDEO RECORDING

Ampex Corp.
401 Broadway
Redwood City, Calif. 94063

Trans-A-File Systems Corp.
371 Santa Trinita
Sunnyvale, Calif. 94086

and others

MISCELLANEOUS CARD SELECTORS AND MECHANIZED RETRIEVAL DEVICES

Access Corp.
4632 Paddock Rd.
Cincinnati, Ohio 45229

BCD Computing Corp.
P. O. Box 240
Buffalo, N.Y. 14225

Dakota Microfilm
9655 W. Colfax Ave.
Denver, Colo. 80215

DASA Corp.
Graphic Products Div.
15 Stevens St.
Andover, Mass. 01810

DSI Systems, Inc.
1225 Connecticut Ave., NW
Washington, D.C. 20036

"Electrofile"
Acme Visible Records, Inc.
Crozet, Va. 22932

Foto-Mem Inc.
2 Mercer Rd.
Natick, Mass. 01760

Image Systems, Inc.
11244 Playa Ct.
Culver City, Calif. 90230

Microform Data Systems Inc.
Suite 1507-Palo Alto Office Center
Palo Alto, Calif. 94301

Microsystems Inc.
1717 Barnum Ave.
Bridgeport, Conn. 06610

Mohawk Industrial Laboratories, Inc.
1 Ward St.
Vernon, N.Y. 13476

Morgan Information Systems
3197 Park Blvd.
Palo Alto, Calif. 94306

Mosler Co.
1561 Grand Blvd.
Hamilton, Ohio 45012

NCR-Industrial Products Div.
3100 Valleywood Dr.
Dayton, Ohio 45429

Randomatic Data Systems, Inc.
344 West State St.
Trenton, N.J. 08618

Regiscope Corporation of America
7 East 43d St.
New York, N.Y. 10017

Remington Rand
Office Systems Div.
2233 Wisconsin Ave., NW
Washington, D.C. 20007

Sanders Associates, Inc.
95 Canal St.
Nashua, N.H. 63060

Technifax
6200 Kansas Ave., NE
Washington, D.C. 20011

Varian ADCO
470 San Antonio Rd.
Palo Alto, Calif. 94306

and others

CLUE WORD EXTRACT CARD

Prepared in-house—a known user is:

Battelle Memorial Institute
505 King Ave.
Columbus, Ohio 43201

and others

PERMUTED INDEXES

Usually prepared in-house by electronic computers. KWIC (keyword in context) and other standard computer programs available from IBM, RCA, and others.

"Spindex"
Dr. Frank G. Burke
Director, Educational Programs Division (NE)
National Archives and Records Service (GSA)
Washington, D.C. 20408

"Wadex"
Dr. Harold Wooster
Chief, Research and Development
Lister Hill National Center for
Biomedical Communications
National Library of Medicine
8600 Rockville Pike
Bethesda, Md. 20014

and others

COLUMNAR CARD

Can be prepared in-house; also available commercially from:

Dataflow Systems, Inc.
7758 Wisconsin Ave.
Bethesda, Md. 20014

and others

DUAL DICTIONARY

Prepared in-house, usually by electronic computer. Users include:

Battelle Memorial Institute
505 King Ave.
Columbus, Ohio 43201

"Kros-Term"
Engleman and Co. Inc.
2480 16th St., NW
Washington, D.C. 20009

and others

EDGE-NOTCHED CARD

"Keysort"
Litton Automated Business Systems
1700 Wisconsin Ave., NW
Washington, D.C. 20007

and others

OPTICAL COINCIDENCE

Better Ideas, Inc.
210 Wayne Dr.
Cinnamon, N.J. 08077

Carter-Parratt Limited
Iddesleigh House
Caxton St.
London S.W. 1 Works: Sutton & Bath
England

Find-It
P. O. Box 25942
Los Angeles, Calif. 90025

Information Retrieval Inc.
801 Welch Rd.
Palo Alto, Calif. 94304

"Keydex"
Litton Automated Business
1700 Wisconsin Ave., NW
Washington, D.C. 20007

"Termatrex Systems"
REMAC International Corp.
26 North Summit Ave.
Gaithersburg. Md. 20760

Scientific Advances, Inc.
1400 Holly Ave.
Columbus, Ohio 43212

Joshua Stern
Chief, Instrumentation Application Section
Room A-351, Building 225
National Bureau of Standards
Gaithersburg, Md. 20234

and others

PUNCHED CARDS

Friden, Inc.
2100 L St., NW
Washington, D.C. 20037

IBM Corp.
1111 Connecticut Ave., NW
Washington, D.C. 20036

UNIVAC
2121 Wisconsin Ave., NW
Washington, D.C. 20007

and others

COMPUTERS (AND COMPUTER MASS MEMORIES)

Essentially any of the existing digital computers can, with proper programming, be used in an information retrieval system. Many computer manufacturers also offer computer mass memories of various capacities. Since there are so many manufacturers, and this information is readily available from other sources it is not included here.

APPENDIX C

INFORMATION RETRIEVAL RECOMMENDED PRIMERS AND SELECTED RESEARCH SOURCES

Methods of Information Handling, Charles F Bourne. John Wiley and Sons, New York, 1963.

A useful source of information of design of information retrieval systems as well as fairly current reference to the various types of equipment and devices.

Introduction to Information and Storage Retrieval: Tools, Elements, Theories, Joseph Becker and Robert M. Hayes. John Wiley and Sons, New York, 1963.

A general introductory text on information retrieval. Explains the uniterm system, especially the printed dual dictionary index; the termatrex and minimatrex systems; punched cards organized as collator decks; magnetic tape and tape search units.

Information Storage and Retrieval: A State-of-the-Art Report, Lawrence Beral. Auerbach Corporation, Sept. 14, 1964. AD 630 089.

An easy to read comprehensive survey of the many areas of activity which are a part of information retrieval. The biggest contribution made by this report is to bring the subject into clear and organized perspective. None of the activities are analyzed in any depth, however.

Principles of Automated Information Retrieval, William F. Williams. Business Press, Elmhurst, Illinois, 1965.

A general discussion of the field of information retrieval showing many different types of equipment. This book is similar in content to reference No. 1 of this list, but does not go into as much depth.

"Information," Scientific American. Vol. 215, No. 3, Sept., 1966.

This source is an excellent discussion of the application of computers to information processing, manipulation, storage and retrieval. It extends beyond the boundaries of information retrieval to include electronic system logic, software definitions, data communications, time sharing, programmed learning, and other topics currently being discussed under the heading, "Automation." The particular value of this report is to show how information retrieval fits into this broader picture.

Microfilm in Business, Joseph L. Kish, Jr., and James Morris. The Ronald Press, New York, 1966.

A useful work on the application of microfilm to the office.

Journal of the American Society for Information Science (formerly American Documentation) and the Proceedings of the American Society for Information Science, No. 11 to date, American Society for Information Science. Washington, D.C., 1964 to date.

These are the most important publications in the field of Information Retrieval. Any survey of the literature must include some of the articles printed in this magazine. At one time or another, every major topic is discussed at great length and depth. To keep current in the field, this is the one most important reference.

Microfilming Technology: Engineering and Related Fields, Carl Nelson. McGraw-Hill, New York, N.Y., 1965.

A basic guide on microfilm, particularly in the field of engineering data.

Annual Review of Information Science and Technology, Carlos A. Cuadra, Editor. John Wiley & Sons, New York, 1966 to date.

A highly useful source of information on the latest developments in information retrieval sciences.

Information Retrieval Systems, Characteristics, Testing, and Evaluation, F. W. Lancaster. Wiley & Sons, New York, to date.

An analysis of the basic elements and essential features of information retrieval systems with particular attention to testing and evaluation.

OTHER SOURCES OF INFORMATION

1. Most of the magazines writing in the general area of automation contain worthwhile articles on information retrieval. Some of the most interesting to office information retrieval are:

   a. Datamation

   b. Data Processing

   c. Systems

   d. Business Automation

2. Other more technical magazines are:

   a. Special Libraries Journal

   b. Association of Computing Machines Journals

   c. Aslib Journal

3. The two most abundant sources of current information on all aspects of information retrieval are Defense Documentation Center, (DDC,) Cameron Station, Alexandria, Virginia and the National Technical Information Service, Department of Commerce, Springfield, Virginia (NTIS). All agencies of the Federal Government and all private companies under contract to the Department of Defense are eligible to receive free research reports from DDC. There are certain restrictions put upon distribution which vary with the report and the source of the request. However, much of the information is available for the asking. For this reason, DDC is one of the most valuable sources of current information on information retrieval in the United States today.

For those people who are not qualified to receive from DDC, NTIS is available. All unrestricted research reports produced under contract by private companies for the Federal Government are available for a standard cost of $3.00 to the general public. Since most reports on information retrieval are not classified, NTIS provides an extensive selection on the subject to the interested reader. Whereas DDC contains only Defense Department reports, the NTIS receives reports from HEW, NASA, AEC, and many other Government agencies, civilian and military.

4. Manufacturer's literature is an important source of information on systems theory as well as equipment characteristics. The International Business Machines Corporation has a series of pamphlets on information retrieval applications in the office, the hospital, the law, etc.

5. The most valuable source of information on microfilm is the National Microfilm Association, Suite 1101, 8728 Colesville Road, Silver Spring, Maryland 20910. Their publications include the Micro-News Bulletin, Journal of Micrographics (formerly NMA Journal) and the proceedings of their annual meetings.

APPENDIX D

SAMPLE FORMS FOR EVALUATING INFORMATION RETRIEVAL SYSTEM POTENTIAL

Evaluating Information Retrieval System Potential INFORMATION FACILITY	EVALUATOR'S NAME DATE

ORGANIZATION AND FACILITY

NAME AND ADDRESS OF ORGANIZATION OF JURISDICTION	TYPE OF RECORDS ☐ FOLDERS ☐ CARDS ☐ OTHER (Specify) CONTENTS OF RECORDS

TITLE OF INFORMATION FACILITY	NO. (Net) OF EMPLOYEES AT FACILITY	BUILDING AND ROOM NUMBER	PHONE NO.

USAGE DATA (*Estimated manhours spent annually in looking up, searching, extracting or correlating information or data at this facility*)

PRIMARY USERS (Organization & Unit)	JOB TITLE	ANNUAL MANHOURS	PRIMARY USERS (Organization & Unit)	JOB TITLE	ANNUAL MANHOURS

EVALUATION FACTORS	YES OR NO	KEY									
1. Annual Additions Equal or Exceed: (*Circle applicable letter, if any*) a. 25,000 pages, if system is used mainly for storage of written information. b. 1,000,000 characters, if system is used for storage of precise data such as names, numbers, etc. c. 2,500 individual items, if system is used mainly for storage of graphic, pictorial, or other matter not covered above (Explain in remarks).		Y	Y	Y	Y	N	N	N	N	N	–
2. Information will be in continuous use for over 5 years and one man-year or more is being used for looking up, searching, extracting, or correlating information or data at this facility.		Y	N	N	N	Y	Y	N	N	N	–
3. Information will be in continuous use for less than 5 years and two man-years or more are being used for looking up, searching, extracting, or correlating information or data at this facility.		N	Y	Y	N	N	N	Y	Y	N	–
4. Time presently required for looking up, searching, etc., information or data at this facility is mainly attributable to limitations of conventional methods.		–	Y	N	–	Y	N	Y	N	–	–
5. The information maintained at this facility could be readily obtained from other source(s) (*Specify sources and locations under remarks*).		N	N	N	N	N	N	N	N	N	Y
CONCLUSIONS											
A. Modern information retrieval seems a likely possibility.		X	X								
B. Likely that present or improved conventional methods will suffice.				X	X		X		X	X	
C. Likely that present or improved conventional methods will suffice; HOWEVER, also consider modern information retrieval systems (*particularly those which use inexpensive equipment.*)						X		X			
D. Consider discontinuance of either this or other duplicate facility (ies), and if duplication is widespread, also consider possibility of a central information service or facility.											X
E. Other (*Specify and explain - use remarks if additional space is required*).											

REMARKS

INSTRUCTIONS - Prepare one of these Decision Tables for each file station, record collection, index file or other information facility at the installation being surveyed. Where reference is made to user manhours, specify those spent by employees of the facility as well as any spent at the facility by personnel from other organizational units. Answer "YES" or "NO" in the appropriate column opposite the Evaluation Factors to indicate the existing situation.

Compare your overall findings with those in the columns under the KEY. (*A dash indicates that it makes no difference whether the answer to that evaluation factor is Yes or No.*) When you find a column that duplicates your answers, place a check mark at the top of the column (*preferably with a colored pencil*). Follow the appropriate column down into the Conclusions column and circle the appropriate X.

Evaluating Information Retrieval System Potential
USER NEEDS

EVALUATOR'S NAME

DATE

BROAD TYPE OF INFORMATION

ORGANIZATIONAL UNIT	USER'S JOB TITLES *(Exclude personnel assigned to operate information facilities)*	NUMBER	PHYSICAL LOCATION	ESTIMATED ANNUAL MANHOURS	PRIMARY SOURCES OF THIS INFORMATION

EVALUATION FACTORS	YES or NO	KEY							
1. 5% or more of users* total man-hours (minimum 1 man-year) are being spent in looking up, searching, extracting, or correlating information or data. *(*Users include all persons who personally do the looking up, searching, extracting or correlation, EXCEPT those assigned to operate the Information Facilities)*		–	Y	Y	Y	Y	N	N	N
2. Current information facilities are INADEQUATE for one or more of the following reasons: *(Circle any that apply)* *A. Pertinent documents or information are regularly being missed or system produces too much non-relevant material or information.* *B. System can furnish documents, only, whereas users would like to receive only portions thereof or precise data.* *C. System cannot satisfy need for retrieving precise data and correlating it.*		–	Y	Y	N	N	Y	Y	N
3. Much faster retrieval speed is needed than could ever be achieved under present or any other conventional method.		Y	N	N	N	N	N	N	N
4. Time presently spent in looking up, searching, extracting, or correlating information or data is mainly attributable to limitations of conventional methods.		–	Y	N	Y	N	Y	N	–
CONCLUSIONS									
A. Modern information retrieval system seems a likely possibility		X	X		X				
B. Likely that present or improved conventional methods will suffice.						X		X	X
C. Likely that present or improved conventional methods will suffice; HOWEVER, also consider modern information retrieval systems *(Particularly those which use inexpensive tools)*				X			X		
D. Other *(Specify and explain)*									

INCONVENIENT FEATURES *(Features NOT necessarily attributable to limitations of conventional methods. CHECK ANY THAT APPLY.)*	
	DIFFICULT TO OBTAIN ACCESS TO INFORMATION
	USERS PREFER TO SEARCH BUT FIND SYSTEM DIFFICULT TO UNDERSTAND OR USE
	USERS NOT ROUTINELY INFORMED OF NEW INFORMATION PERTAINING TO THEIR WORK
	OTHER *(Specify and explain)*

REMARKS

INSTRUCTIONS - Prepare as many of these Decision Tables as needed to collect data during the course of surveying individual user groups to estimate manhours spent in looking up, searching, extracting, or correlating information or data. Summarize your findings by preparing one Decision Table for each of the broad, similar types of information required at the installation being surveyed.
Enter "YES" or "NO" in the column opposite each of the Evaluation Factors to indicate existing conditions.
Compare your overall findings with those in the columns under "KEY" until you find a set that matches yours - place a check mark at the top of that column *(preferably with a colored pencil)* Follow the selected column down to the "CONCLUSIONS" and circle the appropriate X.

APPENDIX E

SAMPLE DIRECTIVE (AIR FORCE) COVERING DOCUMENT MINIATURIZATION SYSTEMS

DEPARTMENT OF THE AIR FORCE
Headquarters US Air Force
Washington DC 20330

AF REGULATION 12–40

5 March 1971

Documentation

DOCUMENTATION STORAGE AND RETRIEVAL (DS&R) SYSTEMS

This regulation provides general information and instructions on the establishment and use of systems for documentation storage and retrieval and assigns responsibilities for related actions. It applies to all Air Force activities that are responsible for initiating, evaluating, approving, operating, or using a DS&R system. It applies whether the documentation is generated by the using activity or is purchased, leased, or otherwise obtained from some other source.

	Paragraph
Related Directives	1
Documentation Storage and Retrieval Systems Explained	2
System Media	3
Systems Interface	4
Responsibilities of Air Force Activities	5
How to Develop a DS&R System Proposal	6
Evaluation and Approval of Proposed DS&R Systems	7
Operation of a DS&R System	8
AF Form 112, "Documentation Storage and Retrieval System Report," RCS: HAF–G12	9
★Microfilming Histories and Related Documents	10
Change to an Approved DS&R System	11
Cancellation of an Approved System	12
Command Supplements and Other Publications	13
Supply of Form	14

Attachments

	Page
1. General Information and Instructions on Use of Microfilm	7
2. Format for DS&R System Proposal	11
3. Responsibilities and Instructions for Preparing, Submitting, and Reviewing AF Form 112, "Documentation Storage and Retrieval System Report," RCS: HAF–G12	14
★4. Policy and Specifications for Microfilming Historical Materials	15
5. AF Form 112 (For Local Reproduction)	21

Figure

1. Sample NBS Resolution Test Chart	19

1. Related Directives. The following directives also relate to DS&R systems. The provisions of those in b through d must be met when the use of ADP equipment for any system is contemplated.

a. **AFR 12–1**—Explains the Air Force Documentation Management Program and assigns responsibilities for its implementation.

b. **Other regulations in the 4 and 12 series**—Give additional information about the programs and the systems discussed in this regulation.

c. **AFR 300–2**—States policy on design and operation of automated data processing (ADP) equipment.

d. **AFR 300–3**—Tells how to design, develop, and implement automated data sys-

---

Supersedes AFR 12–40, 25 Nov. 1969. (For summary of revised, deleted, or added material, see signature page.)

OPR: DAD
DISTRIBUTION: S

tems that use computers, punch card accounting machines, or any auxiliary or peripherial equipment.

e. **AFM 171–9, chapter 5, 9 May 1966**—Provides specifications with which to justify and obtain ADP equipment.

f. **AFR 400–14**—States Air Force responsibility for the recovery, reclamation, and use of silver contained in expended photographic materials and silver-bearing scrap.

2. Documentation Storage and Retrieval System Explained. Any system that can store, index, select, and retrieve documents (pages of information) is a DS&R system, even though it uses only conventional manual filing and research methods. (For many collections of records, manual techniques are still the most effective and economical.) However, in this regulation the term DS&R will be used *only* to refer to a system that uses a combination of manual, mechanical, electronic, photographic, video, and similar techniques to miniaturize, store, index, select, and retrieve documentation. Technological advances are resulting in continuous development of new or improved systems to perform those functions. Because of the variety of available equipment and devices, and their diverse capabilities, the potential for developing effective DS&R systems is almost limitless.

3. System Media. Currently, microfilm is the most commonly used medium for DS&R systems. Attachment 1 contains general information and instructions on its use. As need is indicated, similar information on other media (see AFR 12–1, attachment 1, 20 May 1969) will be published.

4. Systems Interface. Expanded use of mechanical and automated techniques has increased interface between systems regulated by the AFR 300 series and AFM 171–9 and those covered by this regulation. For example: Devices that initially convert data from magnetic tape to microfilm are in the ADP category regulated by the AFR 300 series and AFM 171–9, but devices that further process that same microfilm (developers, duplicators, readers, etc) are in the DS&R category and are regulated by this regulation. When a system interface is anticipated, managerial responsibilities of the staff offices concerned at all levels of command must be fully coordinated, joint endeavors.

5. Responsibilities of Air Force Activities. DS&R systems are established and operated under the Air Force Documentation Management Program (See AFR 12–1.)

a. **Air Force Documentation Management Officer (DMO).** The Air Force Documentation Storage and Retrieval Office (AF/DAD) carries out the following responsibilities of the Air Force DMO:

(1) Develops and coordinates policies and procedures governing the use of DS&R systems, Air Force-wide.

(2) Negotiates with other agencies, as required by law, executive order, or regulation, on Air Force use of DS&R systems.

(3) Evaluates and approves or disapproves the use of DS&R systems by Air Force activities, based on:

(a) Policies and requirements affecting the creation, maintenance, and disposition of Federal records.

(b) Information furnished in the proposal (soundness of system, selection of equipment, method of accomplishment, economies, etc).

(c) Coordination and comments of activities that have either a functional interest in the documentation or a responsibility to support implementation of the system.

NOTE: Approval of a DS&R system does not constitute approval of funds, manpower, or facilities for the system; those resources must be approved through regular command channels.

b. **Command Documentation Management Officers.** Command DMOs will:

(1) Promote the effective use of mechanized and automated systems that will result in economic savings and improved service and advise about and help develop and implement such systems.

(2) Evaluate, justify, coordinate, and submit proposals—for establishing or chang-

ing DS&R systems—to HQ USAF/DAD, Wash DC 20330 as prescribed by this regulation.

(3) Insure that approval of HQ USAF/DAD is obtained before DS&R systems are installed or procurement of equipment, supplies, or service contracts is initiated (see exception in c below).

(4) Insofar as practicable, centralize equipment and operations to support more than one DS&R system (see AFR 12-42).

(5) Provide for periodic review of approved DS&R systems to determine their continued justification and to insure compliance with this regulation.

(6) Submit DS&R Systems Reports as specified in paragraph 9 and attachment 3; take any required action on matters reported.

c. **Activities Responsible for Research and Development (R&D).** An activity that has been delegated the R&D responsibility may implement experimental DS&R systems under an approved development or test directive without obtaining prior approval of HQ USAF/DAD. However, the activity must furnish the command DMO a copy of the approved directive and a report of the results of the experiment, including the disposition to be made of the equipment being tested or developed.

6. How to Develop a DS&R System Proposal. A mechanized or automated system that stores, indexes, selects, and retrieves documents can be advantageous, *if* a realistic analysis (of document content and nature, equipment characteristics, and users' requirements) proves that the total planned system is effective and economically sound. To develop a DS&R system proposal:

a. **Analyze the Documentation:**

(1) Determine and justify the reason for maintaining the collection, and the purpose that it serves. For example: Consider whether the same information is available from another source; consider whether less frequently used documentation can be segregated and retired to a staging area or to a records center (see AFM 12-50).

(2) Determine the requirements for and methods of updating information in the collection, and analyze the current and contemplated system for disposing (purging) of information as it becomes noncurrent.

(3) Be sure the filing arrangement permits easy access to up-to-date information. Ascertain the frequency of each type of reference (view only, borrow, copy) and identify user requirements.

(4) Study the physical characteristics of the paper documents and determine the practicality of converting them to another form.

(5) Identify and define the specific inadequacies of current procedures. Determine whether those inadequacies can be eliminated through improvements resulting from actions in (1) through (4) *above*.

b. **Plan the Proposed System:**

(1) Contact representatives of various commercial firms to obtain a knowledge of available technology. Determine capabilities, limitations, and costs of equipment to establish and continue techniques being considered for the system. Specific equipment, identified by the manufacturer, make, model, etc, may be selected for planning purposes only, but a commitment must not be made nor implied before approval of the proposal by HQ USAF. (Manufacturers' proposals or recommendations for ADP or PCAM equipment will not be solicited; unsolicited proposals will be forwarded to HQ USAF/ACDC, Wash DC 20330 per AFR 300-2.)

(2) Establish the objectives of the proposed system and, after considering all users of the documentation, plan the total system.

(3) Determine the availability of any suitable Air Force-operated equipment that could be used (shared) in the proposed system.

c. **Prepare and Submit the System Proposal:**

(1) When preparing the DS&R system proposal, follow the format in attachment 2. Prepare the proposal in sufficient copies to provide at least one for the command DMO and two for HQ USAF/DAD. Furnish information that will permit a clear understanding of the present system and the proposed system, and provide a justification for the

change. If appropriate, attach sample copies of documents, studies, flow charts, cost analyses, etc.

(2) If a feasibility (pilot) test is planned for the system, indicate objectives and limitations of the test. (The test period will not exceed 6 months, and equipment will be leased and not purchased for the system during the test period.) Submit copies of any progress and final reports of these test systems to HQ USAF/DAD.

(3) If an interface with ADP is contemplated in the proposed system, prepare a single document proposal, and:

(a) When an automated data system change is proposed, submit in accordance with AFR 300–3 and attach information required for the DS&R system proposal (see format, attachment 2). Send an information copy of the entire proposal to HQ USAF/DAD.

(b) When there is no automated data system change involved, submit equipment specifications in accordance with AFM 171–9 and attach information required for the DS&R system proposal format. Send an information copy of the entire proposal to HQ USAF/DAD.

7. Evaluation and Approval of Proposed DS&R Systems. When evaluating the merits of a proposed DS&R system, the primary factors to consider are the operating and administrative procedures that can be improved now or later, and the relative costs of the proposed system as compared with costs for any alternate procedure. The procedures for evaluating and approving or disapproving a proposed DS&R system are as follows:

a. **Command DMOs** will:

(1) Insure that:

(a) The information in the proposal is accurate and complete, and

(b) The techniques and equipment proposed for the system are practical.

(2) Determine whether any of the operations required for the proposed system could be accomplished by sharing the use of existing approved equipment and facilities. If not, furnish reasons.

(3) Coordinate the proposal with all activities that might be affected by, or interested in, the implementation of the system. (Include any required coordination on funds, manpower, and facilities for the proposed system.)

(4) Send recommended proposals to HQ USAF/DAD, for approval. Include with command recommendations any additional information that justifies or clarifies the proposal.

(5) Return incomplete or unacceptable proposals to the initiating activity with appropriate explanations.

b. **HQ USAF/DAD** will:

(1) Approve or disapprove proposed DS&R systems (see paragraph 5a(3)).

(2) Assign a systems control number to each approved proposal (for future identification of the system) and furnish any appropriate contingencies or instructions on establishing and operating the system.

(3) Return disapproved proposals to the command DMO, with appropriate explanations.

8. Operation of a DS&R System. Established systems must be periodically monitored to insure that continuance is justified and that procedures and equipment are improved in keeping with technological advancements. As a minimum, provisions must be made for:

a. **Safeguarding Classified Information:** The same security requirements that apply to safeguarding, storing, shipping, and granting access to classified papers also apply to classified information in a DS&R system. In addition, take these precautions:

(1) Use the services available from Air Force facilities as far as practicable; use commercial firms only when they have been cleared and authorized to handle classified information under AFR 205–4.

(2) Comply with AFRs 205–1 and 205–4, which govern the reproduction of classified documents.

(3) Review each classified document before including it in the system to insure that:

(a) All possible downgrading and declassifying actions have been completed, and

(b) Downgrading/declassifying notations, as required by AFR 205–2, appear on all documents remaining classified.

b. **Inspections:**

(1) Carefully inspect miniaturized images before disposing of related paper copies. Insure that:

(a) All documents intended to be processed have been included, are properly indexed, are legible, and can be satisfactorily retrieved.

(b) System medium meets required standards of quality. (For microfilm, see attachment 1.)

(2) Periodically inspect entire collection to determine whether there is any deterioration from any cause.

c. **Disposition of Documentation.** Comply with disposition instructions specified in the approved system for:

(1) Paper copies of the documents committed to the system.

(2) Miniaturized copies of documents. (Attachment 1 discusses the retirement of microfilm copies of documents that have a permanent retention value.)

d. **Housekeeping and Film Handling Procedures.** Insure that proper housekeeping and film handling procedures prevail, especially in the filming, processing, film duplicating, and cartridge or jacket-loading operations. (See attachment 1, paragraphs 6 and 7, for details.)

e. **System Documentation.** Maintain statistical and other information on a continuing basis. (See paragraph 9 and attachment 3 for reporting requirements.)

9. AF Form 112, "Documentation Storage and Retrieval System Report," RCS: HAF-G12. Reports on each DS&R system must be submitted through documentation management channels on AF Form 112, prepared and submitted as shown in attachment 3. Currently, the report is designed for systems that use microfilm as the documentation medium; when specifically required by HQ USAF/DAD reports will be submitted on DS&R systems that use other media.

10. Microfilming Histories and Related Documents. Major commands that desire to microfilm their histories and related documents (see paragraph 2, attachment 4) must submit the DS&R system proposal required by paragraph 6. In the letter of transmittal, indicate the levels of command for which microfilming of historical documents is being proposed. When HQ USAF/DAD approves the proposed DS&R system, microfilm the histories and related documents as explained in attachment 4.

11. Change to an Approved DS&R System. Approval of a DS&R system applies to the system as initially authorized. When any major change (see a through f below) is contemplated, advance approval of HQ USAF/DAD must be obtained. Proposed changes are to be prepared and submitted in the same manner as the initial proposal. For example, major changes are proposals to:

a. Discontinue inservice operations and contract for the services with a commercial firm, or vice versa.

b. Take over an approved system from another activity.

c. Change, add to, or discontinue any documentation or indexes approved for the system.

d. Change the disposition criteria for any documentation in the system.

e. Change the format of the converted documentation.

f. Obtain equipment, or equipment modifications, not originally approved for the system.

12. Cancellation of an Approved System. To cancel an approved DS&R system, notify HQ USAF/DAD by letter. Include:

a. Date of cancellation.

b. Reasons for cancellation.

c. Disposition of the documentation.

d. Disposition of the equipment approved for the system.

13. Command Supplements and Other Publications. Commands may not issue supplements or other publications that change the basic policies, procedures, or formats prescribed in this regulation. Forward a copy of each MAJCOM supplement or other publication on this subject to HQ USAF/DAD as required by AFM 5–1, paragraph 10–7 and rule 4, table 14–1, 1 October 1968.

14. Supply of Form. Locally produce AF Form 112, on 8" X 10½" paper, as shown in attachment 5.

BY ORDER OF THE SECRETARY OF THE AIR FORCE

OFFICIAL

JOHN D. RYAN, *General, USAF*
Chief of Staff

DWIGHT W. COVELL, *Colonel, USAF*
Director of Administration

Summary of Revised, Deleted, or Added Material

This revision adds procedures for microfilming history and related documents (para 10 and atch 4); substitutes "ANSI" for "USASI" (paras 2d and 6a(2) of atch 1), new office symbols for old, and a new reports control symbol for AF Form 112; and revises instructions for microfilming permanent documentation (atch 1, para 4).

1222

GENERAL INFORMATION AND INSTRUCTIONS ON USE OF MICROFILM

1. Microfilming Explained. Microfilming, or microphotography, is a process by which miniature photographic images of documents are reproduced on film. Generally, these images must be magnified to be read. As technological advances are made in microfilming systems and equipment, various forms of microfilm are being developed and its uses are being expanded.

a. **Microforms.** The more common forms of microfilm are:

(1) *Roll Microfilm.* A length, usually 100 feet, of microfilm that is kept on a reel or in a cartridge or similar type container.

(2) *Aperture Card.* A card with one or more holes, or windows, designed to hold a frame of microfilm. Although the card may be of any size, this term generally refers to tabulating cards that can be mechanically sorted, filed, and extracted by punch card accounting machines.

(3) *Strip Microfilm.* A unit length of microfilm that is too short to be wound on a reel. The strips are inserted in a jacket, stripped on to a sheet of film, or stored in a container.

(4) *Microfiche.* A sheet of film containing multiple microimages, generally arranged in a grid pattern.

(5) *Chip Microfilm.* A unit of microfilm containing one or more microimages and an area for recording code bits.

b. **Sizes.** The width of microfilm is measured in millimeters. The most commonly used sizes are 16mm and 35mm for roll microfilm and 105mm for microfiche. Wider or narrower film generally requires special cameras and readers. The length of microfilm is usually measured in linear feet or inches.

c. **Copies (Prints).** The relation of a copy to the original document is called "generation." The original exposed and developed microfilm is the first generation microfilm copy. Copies made from the first generation microfilm, whether reproduced on film or paper, are second generation prints. Copies made from the second generation prints are third generation prints, etc. Any generation print may be either a positive or a negative, depending on the type of process and the film used. (A positive is identical to the original document in that dark portions appear dark and light portions appear light. A negative is just the reverse.)

d. **Types.** The three most common types of microfilm are: Silver, diazo, and vesicular. For many years, silver film was the only type able to receive the image from the camera. However, recent developments in the use of vesicular film for this purpose have been successful. Although silver film can be used to make prints, either diazo or vesicular film is generally used for the "work copy."

2. Standards and Specifications. The following specifications and standards apply to the various areas of microfilm. Experience proves that compliance with the requirements of these specifications and standards where possible guarantees microfilm of excellent quality.

a. **Federal Standards:**

(1) Fed-Std 125a Film, Photographic and Film, Photographic Processed (For Permanent Record use.)

(2) COSATI Standard PB 167-630, Microfiche.

b. **Military Specifications:**

(1) MIL-M-9868D — Microfilming of Engineering Documents, 35mm; requirements for.

(2) MIL-P-9879A—Photographing of Construction/Architectural Drawings, maps and related documents, 105mm; requirements for.

(3) MIL-M-38748A Microfiche; for Engineering/Technical Data, reports, studies and related data, requirements for.

(4) MIL-M-38761—Microfilming and Photographing of Engineering/Technical

Data and Related Documents: PCAM Card Preparation, Engineering Data Micro-Reproduction System, General Requirements for, Preparation of.

c. **Federal Specifications:**

(1) L-F-315b—Film, Direct Positive, Roll (Diazotype)

(2) L-F-320b Film, Thermal Developing

(3) L-F-334d Film, Photographic, Roll, Microfilm (Black & White)

d. **ANSI Specifications:**

(1) PH 1.28-1957—Photographic Film for Archival Records.

(2) PH 1.29-1958—Curl of Photographic Film, Methods for Determining the.

(3) PH 2.19-1959—Diffuse Transmission Density

(4) PH 4.8-1958—Determining Thiosulphate Content of Processed Black and White Photographic Film and Plates; Method for. (There is a small fee for ANSI items.)

3. **Legal Status of Microfilmed Records.** A record's legality and admissibility as evidence in court are not affected by the fact that the record has been microfilmed. A microfilmed copy of a record is admissible in evidence in a legal action when it can be established that the record was microfilmed by the authorized custodian of the records as a routine, controlled procedure approved by higher authority. The same substantiating measures also are sufficient to establish the authenticity or legality of microfilmed records for audit or investigative purposes (44 U.S.C. 399).

★4. **Microfilming Permanent Documentation:**

a. **Authorized Production.** When documentation that has a permanent retention value is microfilmed, only the following production is authorized, unless HQ USAF/DAD specifically approves others. Care must be taken to preclude scratching the master films (camera negative and silver positive).

(1) One camera master silver negative, and produced from it:

(a) One silver positive print.

(b) One diazo copy for internal use or further reproduction purposes.

(2) One silver negative film reproduced from the silver positive film.

b. **Copies for National Archives.** Retire to the National Archives or to the appropriate records center the camera master silver negative and the positive silver print made from it (see AFM 12–50) after inspecting them to assure that they are adequate substitutes for the original documents (see FED-STD 125 and ANSI Specification PH-1.28-1957).

5. **Microfilm Equipment.** Basic equipment for a microform system consists of a camera, a processor, and a viewer. In addition, duplicators, viewer-printers, retrieval keyboards, inspection kits, splicers, and a variety of other accessories and peripheral devices are available. It is not feasible to identify and describe the function of all types of microform equipment; however, some general information and guidance are provided below.

a. **Description of Microform Equipment:**

(1) *Microfilm Cameras.* Microfilm cameras are referred to as being either planetary or rotary cameras. When a planetary camera is used, both the film and the document are still; when a rotary camera is used, both the film and the document are in motion. Most rotary cameras use 16mm film, and most planetary cameras use 35mm film; however, there are planetary and rotary cameras that will accept either 16mm or 35mm film. A step-and-repeat camera is a planetary type camera used for producing microfiche. Using 105mm film, it microfilms an image at a time, completes a row of images, then returns and repeats the process on the next row in the microfiche grid.

(2) *Processors.* The microfilm processor accepts the roll of film that has been exposed in the camera, develops the latent images, washes the chemicals from the film, and dries and rewinds the film onto a reel. Variances in processors include factors such as speed, daylight or darkroom loading, and film sizes (width and length). Processing exposed microfilm is perhaps the least time-consuming operation of any system; one microfilm pro-

cessor is capable of supporting several systems.

(3) *Viewers and Viewer-Printers.* Equipment for viewing or reading microfilm images differs widely. Some can print a copy of the document being read (viewer-printers); others (viewers) cannot. (Usually the print copy approximates the size and quality of the document that was originally microfilmed.) Some will accommodate only one form of microfilm (roll, microfiche, aperture card, etc); others will accept combinations of microforms. Since this equipment is the retrieval station, many different devices and ways to select the desired images are available with it.

(4) *Other Equipment, Acessories, and Devices.* Obtain brochures from, and discuss other types of microform equipment with, your local manufacturers' representatives.

b. **Selecting Microform Equipment.** System objectives and requirements dictate parameters for selecting equipment, just as a knowledge of equipment capabilities and limitations may affect the system design. The comparison of cost versus worth dominates most determinations. The local representative's reliability for service and maintenance must be considered when selecting equipment.

c. **Procuring Microform Equipment.** Equipment identified in an approved DS&R system is authorized unter TA 006, section D. Before initiating procurement actions for this equipment, consider:

(1) Sharing use of existing Air Force operated equipment.

(2) Lease of equipment with option to purchase, particularly for short-term use and for systems that are subject to changes that may require corresponding changes to equipment. In these instances, a service contract to accomplish the filming and/or processing may be more economical than either purchasing or leasing the equipment.

(3) Maintenance contracts on purchased equipment.

6. Checklist for Contract or In-House Microfilm Operation. This list is not all inclusive, nor is it intended to infringe upon the many laws and regulations on procurement and contract operations. Experience indicates that the several important items and procedures necessary to obtain desired results are:

a. **A visit to the microfilm facility or laboratory** to:

(1) Examine the facility's physical condition, manpower, and capability to complete the proposed job in the desired manner (for example, skilled technician, plant capacity, etc).

(2) Observe the chemical mix area. This area should provide clean mixes and have properly mixed chemicals that comply with ANSI standards and preclude oxidation of the chemicals. Check the procedures in this area. Ask if there is a copy of the current ANSI standards on file. If there is, review to insure currency.

(3) Discuss the facility's record-keeping practice. Explain any requirements the vendor or technician does not understand (inspection sheets for processed film, indexes, processor quality control sheets, etc).

(4) Check the quality control program and procedures in the processing area; for example, the control over processing solution temperature, processing time, control over replenishment-chemicals flow into the processing equipment, etc.

(5) Evaluate the film inspection facilities and procedures. It is most important that the resolution, density, residual hypo, etc, be checked on all film processed, and that the instruments for these inspections be capable of performing according to existing specifications. Observe, also, whether a frame-by-frame check is performed on the processed film passing through the laboratory.

(6) Verify the technical competence of the personnel assigned, and note whether manning is adequate for satisfactory performance.

(7) Inspect the overall cleanliness and the film handling techniques. Do personnel wear white cotton gloves at all tims when handling film? This requirement applies to all operations that involve film handling.

Attachment 1

(8) Observe the procedures involved in camera operation. Do the camera operators handle submitted data carefully? Do they display technical competence in their jobs? Do they wear dark or other nonreflecting garments while operating the cameras?

b. **A sampling of the contractor's work.** If possible, before contract award, submit excellent, medium, and poor-quality samples to each vendor being considered. Inspect his efforts when he returns the samples to see whether:

(1) His performance is satisfactory, and

(2) Poor quality material will microfilm satisfactorily.

7. Care in Handling Microfilm. To protect microfilm from scratches, lint, dust, and other materials that might damage or distort the film image, take the following precautions:

a. Always wear white-cotton, lint-free gloves when handling film of any kind.

b. Always hold film by its edges, never on its flat surfaces.

c. Never wind film on the reel too tightly nor grab or hold the end of the film and pull it to tighten it on the reel. Wind film on the reel only as tight as the camera, viewer, or other rewind mechanism permits.

d. Before using viewers, microscopes, densitometers, etc, always clean all parts that will come in contact with the films, and keep them clean during use. Dust is a deadly enemy to film.

e. After the service (workhorse or use) copy of the microfilm has been inspected, accepted, and produced, use the original camera negative only in emergencies.

f. Never use a viewer that has revolving flats to view camera negative microfilm or any other microfilm. Such viewers tend to scratch the film (especially in the diagonal plane of the film).

g. Keep chemicals and chemical fumes away from film-storage and film-use areas. Never use alcohol to clean film; use film cleaner (photographic).

h. Wind and store film on plastic instead of metal reels. In humid areas metal tends to rust and the rust penetrates the film emulsion.

i. Periodically check stored film. This check is most important. Check for fungus, mildew, film stickiness, too tight winding, image etching, etc. Perform this check at least annually and more often in areas of high or very low natural humidity, especially if film is not maintained under controlled temperature and humidity.

8. GSA Microfilming Services. The GSA regional offices offer Federal agencies the services described below. Consider them when developing a microform-based DS&R system.

a. **Central Source of Information on Microfilming.** This service includes furnishing information on current uses of microforms and on new techniques and developments in this field.

b. **Technical Advice and Assistance.** This service is designed to promote programs to:

(1) Preserve records;

(2) Reduce volume;

(3) Provide security copies;

(4) Make duplicate copies; or

(5) Improve information retrieval system.

c. **Central Reimbursable Microfilming Service.** This service includes preparing, indexing and filming records; inspecting film; and labeling film containers.

Attachment 1

FORMAT FOR DS&R SYSTEM PROPOSAL

(Type on Appropriate Letterhead)

REPLY TO
ATTN OF: (Office symbol of requesting office)

SUBJECT: Proposed DS&R System for (Brief title of documentation and location)

TO: (Major command documentation management officer)

1. Name of activity. (Show complete organizational identification and location where proposed system will be established.)

2. Present system:

a. Document description. (Indicate records series by title and appropriate table and rule number from AFM 12-50. When more than one series is involved, list each series. Show inclusive dates of the documentation, its security classification, and the physical characteristics of the documents (paper, microform, or other media). When other than paper documents are involved, state whether original paper documents are available, or whether the medium used is a suitable substitute for the paper records. Show largest, smallest, and majority of document page sizes; estimate number of documents in the files and average number of pages in each document. Include any other remarks pertinent to description of the documentation.)

b. File volume. (State on-hand volume in cubic feet; estimated weekly, monthly, or yearly accumulation in cubic feet. If more than one series of documents, show volume for each series.)

c. File (document) maintenance and disposition. (Describe kind of files, and whether centralized or decentralized. Describe filing arrangement, related indexes, method of updating and other changes, number of updates or other changes per day, week, or other, and system of disposing of noncurrent information. Discuss indexing and filing procedures.)

d. Retrieving and furnishing information. (Describe method of referencing the file, and show the number of references per day, month, etc. Identify duplicates and related records series and their locations. Identify typical users of file by office, position type (engineer, scientist, manager, etc), and grade ranges. List sample questions asked, time required for typical search, percentage of searches that require copies of documents to be made, and method of copying).

e. Manpower. (List Personnel (by grade and AFSC) presently employed in indexing, filing, retrieving and copying documents. Show manhours spent in filing, updating, indexing, retrieving, copying, and disposing of documents. Explain whether personnel are full or part time, military or civilian.)

f. Equipment and floor space. (List equipment used for maintaining document files. Include all file cabinets, sorting racks, desks, tables, bookcases, etc, directly related to file operation, and indicate floor space required for operation.)

Attachment 2

3. Proposed system:

a. System description. (Describe proposed storage medium—aperture card, microfiche, jacket, roll film, video tape, etc—and include microform size and reduction ratio to be used. Explain procedures for converting files and implementing the system; indicate number of copies and distribution of film to be made; discuss method of retrieval, how film will be duplicated, etc.)

b. File volume. (Show cubic feet of on-hand holdings and estimated number of images to be converted to new system. Show estimated annual volume to be entered into system.)

c. Files maintenance and disposition. (Describe in detail all changes that will result from conversion to proposed system. For example: Will documentation be centralized or decentralized? What changes will be made to filing arrangement, indexes, method and frequency of updating or other changes to information in file? If a thesaurus or dictionary is to be developed, what is the status? Indicate method of disposing of non-current information from converted documentation and include disposition of paper records after they have been converted, and disposition of the converted documentation.)

d. Retrieving and furnishing information. (Explain method of referencing, estimate number of references per day and average time for search, and indicate type of questions to be asked of file, if different from paragraph 2d. Identify any difference in users described in paragraph 2d above. Explain method of furnishing information to searcher and the percentage of searches that will result in duplicate microform copies or hard copy enlargements.)

e. Manpower. (Estimate, by grade and AFSC, the manpower required to operate the proposed system on a continuing basis and the additional temporary personnel that will be required to establish the system and convert the backlog. Show manhours required for each operation (microfilming, inspecting, mounting, etc). If presently assigned personnel are to be used, explain the required training and its costs. If contractor service is to be used, explain actions taken under AFR 26–12.)

f. Equipment. (Identify each item of equipment required for the proposed system; furnish make and type, if known, and reasons for selecting a particular equipment. Identify any Air Force-operated equipment available for use in the proposed system, or explain reasons for not sharing the use of existing equipment. Indicate whether rental or purchase is contemplated, and the cost of each item. Explain disposition to be made of equipment used for present system; see paragraph 2f above.)

g. Supplies. (Itemize supplies and costs contemplated to implement proposed system, and estimate annual requirements.)

h. Facilities. (Explain increase or decrease of floor space for proposed system over current system, after backlog is converted. Describe any changes in facilities or utilities required for new system and furnish costs.)

i. Resources. (Provide information about approval of funds, manpower, and facilities for the proposed system.)

Attachment 2

j. Schedule. (Prepare a schedule showing significant events and their estimated beginning and ending dates; for example: Procuring and installing equipment, modifying facilities, beginning of indexing, beginning and ending dates for backlog, starting dates for conversion of current documents. For scheduling purposes, the day the systems approval is received will be "O" day. When slippage occurs, send schedule of adjustments and reasons for slippage to the command documentation management officer.)

4. Justification. (Explain advantages of new system over old, cost reductions, management improvements, etc.)

FOR THE COMMANDER (or other appropriate closing)

(No.) Attachments
(Attach sample documents, flowcharts, cost studies, etc, as appropriate.)

Attachment 2

RESPONSIBILITIES AND INSTRUCTIONS FOR PREPARING, SUBMITTING, AND REVIEWING AF FORM 112, "DOCUMENTATION STORAGE AND RETRIEVAL SYSTEM REPORT," RCS: HAF–G12

1. Responsibility for Preparing Report. Each activity that is responsible for producing microfilm (either original camera microfilm or duplicate film) will prepare AF Form 112, "Documentation Storage and Retrieval System Report." This requirement includes activities that are responsible for service contracts for microfilming, processing, duplicating, or similar microfilm operations. It does not apply to leased commercially prepared microfilm, such as the VSMF, Showcase, and similar systems.

2. Report Preparation. The report is self-explanatory; if exact information is not known enter estimated statistics and indicate that they are estimates. Use additional sheets of blank paper if needed. Prepare a separate report for each separate DS&R system. In addition, if it meets the requirements in paragraph 1:

a. **Each Engineering Data Service Center (EDSC)** will prepare a separate report, even though it operates under a common DS&R system number (AFLC 1–58).

b. **Each Accounting and Finance Office (AFO)** will prepare a separate report, even though the office operates under two common DS&R system numbers (AFAFC 1B-54 and AFAFC-2B-69).

c. **Each Base Civil Engineer** will prepare a separate report, even though it operates under a common DS&R System number AF/PRE 2B-63.

3. Submission of Reports and Responsibilities for Review:

a. **The preparing office** will submit three copies of the report to its documentation manager (DM). (See AFR 12–1 for assigned responsibilities.)

b. **The Activity's DM** will:

(1) Review the report for completeness and assist the preparing office, as required.

(2) Retain one copy of the report for followup action and information until it is replaced by the next report.

(3) Send the original and duplicate to the command DMO. (Exception: Send reports from AFOs to AFAFC/SUAD, 3800 York St, Denver CO 80205, and a courtesy copy to the DM of the host activity.)

c. **The command DMO** will:

(1) Insure that all reports have been received from intracommand activities required to submit reports.

(2) Submit by transmittal letter to HQ USAF/DAD the original of each report recived; include in the letter any appropriate comments concerning the report.

(3) Retain duplicates of reports for followup action and information until they are replaced by the next report or are no longer needed for managerial purposes.

4. Reporting Periods and Due Dates. Submit reports annually to cover the period 1 June through 31 May. Reports are due in HQ USAF/DAD by 31 July.

Attachment 3

★POLICY AND SPECIFICATIONS FOR MICROFILMING HISTORICAL MATERIALS

1. Application of Policy and Specifications. The policy and specifications in this attachment apply to all major commands and their subordinate activities that microfilm documents for historical purposes.

a. Command historical offices are encouraged to establish microfilming projects as required. When a project is being established, the command historian should coordinate his initial plan with the Historical Research Division, Air University (ASI/HO) to insure that it is compatible with the activities of that headquarters. Allow at least 30 days for ASI/HO's reply. When reply is received, the command DMO will process the plan as a DS&R system under this regulation. Allow 30 days for final determination of approval by HQ USAF/DAD.

b. The specifications outlined in this attachment for microfilming historical records and related documents are essential to meet quality requirements of the National Archives and Records Service; the Office of Air Force History (AF/CHO), HQ USAF; Historical Research Disivion, AU; the Documentation Systems Division (DAD), HQ USAF; and other Air Force activities. (As used in this attachment, the term "historical records and related documents" includes all unit histories, monographs, special studies, and CHECO reports, with associated supporting documents and other historical material)

2. Division of Effort. Since command historical archives contain some documents, particularly unit histories, that duplicate those deposited in the archives at Maxwell, microfilming activities must be divided to minimize duplication.

a. **Unit Histories.** Although some exceptions may be made by agreement between ASI/HO and a command historical office, the filming of unit histories and their supporting documents will usually be divided as follows:

(1) ASI/HO will microfilm unit histories *below* numbered Air Force (or comparable) level, and will produce copies of film when requested by Air Force commands.

(2) Any command historical office that establishes a microfilming project may microfilm its own histories and the unit histories of its assigned numbered air forces (or comparable levels). Before beginning any microfilming effort, the command historical office must coordinate with ASI/HO to insure that the histories have not already been microfilmed elsewhere.

b. **Other Documents in Command Archives.** The command must coordinate with ASI/HO to prevent possible duplication and arrange for exchanges of film whenever it appears that the command's archives contain significant duplication of documents held by the Air Force historical archives. Such coordination is necessary, for example, before microfilming command monographs and other documents, if the permanent record copies have been deposited in the Air Force historical archives.

3. Microfilm Film Requirements:

a. The film used to microfilm history and related documents will be as follows:

(1) 16mm nonperforated of the anti-halation undercoat (AHU) type, or equal.

(2) Microfilm roll length—not more than 100 feet, including a 6-inch leader and trailer of blank film on each roll.

b. Splicing into a roll of microfilm should be avoided. However, when you must splice film, use the Recordak Presstape Microfilm Splicer, Model 5A, or equal.

4. Camera Equipment. Camera equipment used to film the histories and related documents can be either rotary or planetary. However, it must produce microfilm that meets all requirements in this attachment, and must provide for placing a standard image control mark (blip) on the bottom center of each frame of microfilm.

5. Specifications and Standards. The current issues of the following documents apply:

a. **Specifications—Federal:**

(1) L-F-315b—Film, Direct Positive, Roll (Diazotype).

(2) L-F-320b Film, Thermal Developing.

(3) L-F-334d Film, Photographic Roll, Microfilm.

(4) PP-B-636—Box, Fiberboard.

b. **Standards**

(1) *Federal*: FED-STD-125a Film, Photographic and Film, Photographic Processed (For Permanent Record Use).

(2) *Military*:

(a) MIL-STD-105—Sampling Procedures and Procedures and Table for Inspection by Attributes.

(b) MIL-STD-129—Marking for Shipment and Storage (Federal Government activities may obtain copies of Federal Specifications and Standards, and the Index of Federal Specifications and Standards from established engineering data service centers (see AFR 12–41)).

c. **Other Publications**

(1) ANSI-PH-1.28-1957—Photographic Film for Archival Records

(2) ANSI-PH-2.19-1959—Diffuse Transmission Density

(3) ANSI-PH-4.8-1958—Determining Thiosulphate Content of Processed Black and White Photographic Film and Plates; method for

(4) ANSI-PH-5.6-1961—100 foot reels for processed 16mm and 35mm microfilm, dimensions for (American National Standards Institute (ANSI) standards may be obtained from 10 E. 40th St, New York NY 10016) (There is a minimal charge for ANSI publications.)

6. Quality. The legibility and archival quality criteria for microfilm are as follows:

a. **Acceptable Reduction Ratios:**

(1) Minimum—24x

(2) Maximum—26x

b. **Density.** The background density of processed microfilm, as measured by the method prescribed in ANSI-PH-2.19-1959, will be 0.95 to 1.50 for the camera master negative. When that density cannot be attained because of contrast differences within the documents caused by age, erasures, etc, a density as close as possible to it may be used if legible hard (paper) copy can be produced by a viewer-printer from third-generation negative microfilm.

c. **Resolution.** A minimum resolution of 100 lines/mm is required for the camera master negative. Determine resolution by exposing the NBS Microcopy Resolution Test Charts (see figure 1) as the first and last exposure on each roll of microfilm produced. To measure resolution, use a microscope having a minimum magnification ratio of 50x to measure the exposed targets. Measure the patterns in figure 1 in all 5 positions for planetary equipment and in the center only for rotary equipment.

d. **Archival Quality.** The residual hypo content and archival quality of the camera negative and of the silver positive print film made from it must meet the requirements of FED-STD-125 and ANSI PH-4.8-1958.

7. Number of Copies and Distribution. The number of copies of microfilm and the distribution will be as follows:

a. The original camera master negative and the silver positive print made from it will be forwarded to the National Archives, Wash DC 20408, in accordance with provisions of AFM 12–50.

NOTE: The camera master negative will only be used to reproduce the silver positive print and/or a "workhorse" diazo copy. All other copies required will be reproduced from the silver positive or "work horse" diazo film.

b. Two duplicate silver negative copies will be forwarded to ASI/HOA Building 914, Maxwell AFB AL 36112. ASI/HO will forward one copy to AF/CHO.

c. Other copies as needed will be made by the major command for internal headquarters use or for distribution to subordinate activities.

NOTE: When the Miracode System is used and a requirement exists for reproducing the code patterns for use to automatically retrieve images, all reproductions must be silver-to-silver or the code patterns will not function correctly.

Attachment 4

8. Quality Control. A quality inspection must be performed on all microfilm produced. To preclude unnecessary damage to the original camera master negative, comply with paragraph 8b, basic regulation; paragraph 6 and 7, attachment 1; and the following:

a. **Master Camera Negative.** Perform *only* resolution, density, and residual hypo content inspections.

b. **Duplicate Silver Negative or Diazo Negative.** Perform frame-by-frame check for legibility.

NOTE: If the duplicate silver negative or the diazo negative is not legible, the camera master negative can be checked to ascertain if original filming quality was poor or if reproduction equipment is causing poor quality.

9. Arrangement of Material on Microfilm. To obtain as much uniformity as possible in page arrangement, comply with the following:

a. **Page position for microfilming:**

(1) *Standard-size page.* Film in position to be readable, without modification, on a reader or reader printer with a 13″ X 13″ viewing screen that produces 8½″ X 11½″ prints; that is, film with the top and bottom of each page parallel to the edges of the microfilm.

(2) *Oversize pages.* Photograph oversize pages (those that require multiple frame exposures for 16mm film at 24x to 26x reduction) in sections from left to right and then from top to bottom, with a minimum of 1-inch overlap between adjacent sections.

b. **Sequence (see paragraph 10a(1)) below):**

(1) *Required Sequence.* Film the complete series of histories of one unit (command or numbered air force) in sequence. Keep the volumes of any periodic installment in order, and film successive installments in sequence by date from earliest date to the most recent. Leave at least 2 inches of blank film between each history and between each volume.

(2) *Complete Series.* Do not split a series by filming some installments and omitting others, for example, by filming narratives and omitting supporting documents.

(3) *Divisions of Unit History Films.* Do not divide any installment of a unit history between two rolls of film except between volumes that originally (in the first copy) were bound separately; do not divide narratives and supporting documents between rolls if they were bound together in the original paper copy. On a rare occasion, an exception may be necessary because of the great length of the history being microfilmed. When it is, break the film at a logical place, such as the beginning of a new chapter of the narrative or a new section of the supporting documents.

(4) *One History Per Roll.* Do not put histories of more than one unit on any roll of film sent to ASI/HO.

10. Security Classification. The security classification of a completed microfilm roll must correspond with the highest classification marking on any document in it. If a roll contains no classified documents, show "UNCLASSIFIED" on the NBS Microcopy Resoluton Test chart (see figure 1). If a roll contains classified documents, show its security classification.

a. **Separation by classification.**

(1) Microfilm Top Secret (TS) histories, or TS volumes or annexes of histories, on separate TS rolls. (Paragraph 9b is modified to the extent necessary to provide for separate filming of TS materials.)

(2) Microfilm together, as indicated in paragraph 9b, histories that are classified Secret or lower and those that are unclassified.

b. **Roll and Copy Numbering.** For accounting and control purposes, the rolls, as well as the copies, must be numbered (see *c below*). A simple method would be to number the rolls in sequence, with the roll number followed by the copy number (for example, 10/1 for Roll 10, Copy 1; 10/2 for Roll 10, Copy 2; etc), but each command may devise its own system.

c. **Labeling:**

(1) Comply with AFRs 205–1, 205–2, and 12–31 and with other applicable USAF directives when marking reels, boxes, etc for security and other restrictions on use.

(2) Put roll and copy number on each reel, box, etc (see b *above*).

(3) Do not put any other markings on the reels and boxes sent to ASI/HOA.

d. **Shipping Inventory.** Include a complete inventory of the contents with each copy of each roll distributed as indicated in paragraph 7. In the inventory, list the contents in the order in which the documents appear on the roll of microfilm and include the following information:

(1) Information relating to the roll:

(a) Roll number.

(b) Overall classification plus any other restrictions on use of the roll.

(2) Information relating to each document on the roll:

(a) Title.

(b) Date(s).

(c) Volume numbers.

(d) Any other descriptive information necessary for quick and positive identification of the document.

(e) Image number for the beginning of the document.

(f) Security classification of the document.

11. Contingency Operations. Exceptions to the preceding instructions in this attachment may be required for contingency operations; for example, the reduction factor (paragraph 6a) may have to be reduced to 22x to permit use of portable cameras for microfilming in a combat theater. Exceptions will be authorized by special instructions issued by HQ USAF/DAD as required.

Attachment 4

DOCUMENTATION STORAGE AND RETRIEVAL SYSTEM REPORT *(If more space is needed use remarks and continue on an 8 x 10½" sheet)*	1. REPORT PERIOD FROM	THRU	*REPORTS CONTROL SYMBOL*
THRU:	**TO:**	**FROM:** *(Name and Location of Organization or Staff Agency)*	
2. DS&R SYSTEM NUMBER	3. NO. OF MILITARY & CIVILIAN PERSONNEL ASSIGNED TO DS&R SYSTEM		
	GRADE	AFSC	FULL TIME / PART TIME

I. PRODUCTION AND DISPOSITION

4. MICROFILMING OPERATION

TYPE CAMERA USED	NUMBER OF FRAMES BY SIZE 16MM	35MM	OTHER *(Specify)*
ROTARY			
OVERHEAD			
STEP AND REPEAT			

5. NUMBER OF 100' ROLLS OF FILM			6. UNITIZING *(Number of Frames)*			
BY MM SIZE	PROCESSED	LOADED IN CARTRIDGES	MOUNTED IN APERTURES	INSERTED IN JACKETS	STRIPPED-UP FICHE	OTHER *(Specify)*

7. DUPLICATING *(Number)*				8. DISPOSITION *(Number)*		
	SILVER	DIAZO	VESICULAR	TRANSFERRED *(Distributed)*	RETIRED	DESTROYED
APERTURE CARDS						
MICROFICHE						
100' ROLLS 16MM						
100' ROLLS 35MM						
OTHER *(Specify)*						

II. COST DATA

CONTRACTED SERVICES	COST	INHOUSE	COST
MICROFILMING	$	EQUIPMENT	$
PROCESSING		MAINTENANCE CONTRACT	
LOADING/UNITIZING		SUPPLIES	
DUPLICATING		MANPOWER	
OTHER *(Specify)*		SPACE	
		UTILITIES	
		OTHER *(Specify)*	

9. EXPLANATION OF COST DATA

REMARKS *(Include contemplated changes to system, problem areas, etc.)*

DATE	TYPED NAME AND GRADE OF PREPARING OFFICIAL	SIGNATURE
DATE	TYPED NAME, GRADE & OFC SYMBOL OF DOCUMENTATION MGR	SIGNATURE

AF FORM OCT 69 112

Attachment 5

☆ U. S. GOVERNMENT PRINTING OFFICE: 1971 O—445-594